AF

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant:	Han Htun and Gordon L. Hager	**Examiner:**	Bradley L. Sisson, Ph.D.
Serial No.:	10/001,486	**Group Art Unit:**	1634
Filed:	November 15, 2001	**Docket:**	30426.1USD1
Title:	METHODS FOR SCREENING LIGANDS THAT ACTIVATE THE TRANSLOCATION OF A STEROID RECEPTOR TO THE NUCLEUS IN MAMMALIAN CELLS		

CERTIFICATE UNDER 37 CFR 1.8:
The undersigned hereby certifies that this paper, as described herein, is being deposited in the United States Postal Service, as first class mail, in an envelope addressed to: Mail Stop Appeal Brief – Patents, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on June 8, 2007.

By: _____
Richelle Ann Domingo

55 S. Lake Avenue, Suite 710
Pasadena, California 91101
June 8, 2007

MAIL STOP APPEAL BRIEF - PATENTS
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

We are transmitting herewith the attached:

- ☒ Transmittal Sheet in triplicate containing Certificate of Mailing under 37 CFR §1.8
- ☒ Amended Appeal Brief under 37 C.F.R. §41.37(d)
- ☒ Evidence Appendices 1-4
- ☒ Return postcard

Please charge any additional fees or credit overpayment to Deposit Account No. 50-0306. A duplicate of this sheet is enclosed.

MANDEL & ADRIANO
55 S. Lake Avenue, Suite 710
Pasadena, CA 91101
(626)395-7801

By: _____
Name: Sarah B. Adriano
Reg. No.: 34,470
Customer No. 26,941

Dkt. 30426.1USD1/SBA/TYL

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants	:	Han Htun and Gordon L. Hager	
Serial No.	:	10/001,486	Examiner: Bradley L. Sisson, Ph.D.
Filed	:	November 15, 2001	Group Art Unit: 1634
For	:	METHODS FOR SCREENING LIGANDS THAT ACTIVATE THE TRANSLOCATION OF A STEROID RECEPTOR TO THE NUCLEUS IN MAMMALIAN CELLS	

55 S. Lake Avenue, Suite 710
Pasadena, California 91101
June 8, 2007

MAIL STOP APPEAL BRIEF - PATENTS
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

SIR:

AMENDED APPEAL BRIEF UNDER 37 C.F.R. §41.37(d)

Applicants hereby appeal to the Board of Patent Appeals and Interferences ("the Board") of the United States Patent and Trademark Office ("the Office") from the final Office Action issued May 18, 2006. Applicants filed a response to the Office Action with one-month extension of time on September 18, 2006. Applicants filed a Notice of Appeal on September 18, 2006 with the proper fee. On October 12, 2007, the Office issued an Advisory Action Before the Filing of an Appeal Brief. An Appeal Brief was due November 18, 2006. An Appeal Brief with four month extension of time was filed on March 19, 2007 with the proper fee. On May 21, 2007, the Office issued a Notice of Non-compliant Appeal Brief. The deadline for filing a response to the Notification is now due on June 21, 2007. In response to the Notice, Applicants submit herein an amended Appeal Brief in compliance with 37 C.F.R. §41.37. Accordingly, this amended Appeal Brief is timely filed.

B. Table of Contents

C. REAL PARTY IN INTEREST

Han Htun and Gordon L. Hager are the inventors of the claimed invention. The assignee of record is Han Htun and The Government of the United States of America as represented by The Secretary of the Department of Health and Human Services c/o National Institute of Health.

The real parties of interest are Han Htun and The Government of the United States of America.

D. RELATED APPEALS AND INTERFERENCES

At the present time there are no pending appeals or interferences related to this case.

E. STATUS OF CLAIMS

Claims 1-15 and 19-58 have been cancelled. Claims 16-18 are pending. A copy of the claims is attached in the Claims Appendix.

Claims 16-17 have been rejected by the Examiner and Claim 18 has been allowed.

Claims 16-17 are under appeal.

F. STATUS OF AMENDMENTS

An Amendment was filed on September 18, 2006, in response to the final Office Action dated May 18, 2006.

The Advisory Action Before the Filing of an Appeal Brief dated October 12, 2006, entered the September 18, 2006, Amendment.

G. SUMMARY OF CLAIMED SUBJECT MATTER

Applicants' invention involves:

A method of screening for a ligand that activates the translocation of a steroid receptor to the nucleus in a mammalian cell comprising:

a. contacting a mammalian cell having a nucleus with the ligand, wherein the cell has a plurality of steroid receptor response elements, wherein the steroid receptor response elements comprise a plurality of AGAACA (SEQ ID NO:4) or AGGTCA (SEQ ID NO:5), in an array such that the response element can be directly detected when bound by fluorescently labeled steroid receptor; and

b. detecting the location of fluorescence within the cell,

a change in the relative fluorescence of the nucleus to the cytoplasm so as to increase the fluorescence of the nucleus indicating a ligand that activates the translocation of a steroid receptor to the nucleus in a mammalian cell. (Claim 16 in the Claims Appendix)

Support for the claim may be found in the originally filed specification at page 7, line 20 to page 8, line 15; page 28, line 27 to page 29, line 5; page 29, lines 14-25; page 59, lines 3-7; originally filed Claims 1, 5, 10 and 16; and originally filed Figures 4 and 5.

H. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

The Office rejected Claims 16 and 17 under 35 U.S.C. §112, first paragraph. Specifically, the Office asserts that the specification fails to provide adequate descriptions of: 1) "other cell lines" which can be used for the invention i.e., the specification allegedly does not provide adequate written description for the "genus" of cells (mammalian cells with a plurality of steroid receptor response elements) encompassed by the claims, and 2) how such cells are to be used for the invention.

I. ARGUMENT

Adequate Written Description for a "Genus" of Mammalian Cells with Multiple Steroid Receptor Response Elements

35 U.S.C. §112 requires that, "the specification contain a written description of **the invention**, and ... enable any person skilled in the art to which it pertains, to make and use the **same** ..." (emphasis added). One is not required to enable any more than what is claimed.

The specification complies with 35 U.S.C. §112 because it provides adequate written description for the "genus" of cells required by the claims (i.e., a mammalian cell that has a plurality of steroid response elements) for the following reasons.

First, the specification as originally filed (see Evidence Appendix 1) discloses actual enabling experiments embodying the claimed methods. Specifically, the application discloses cell lines with multiple steroid response elements: 3134 cells (in the specification at page 51, line 21 to page 56, line 7; Figures 1-2); 1471.1 cells (in the specification at page 39, lines 23-30; page 18, lines 4-7; Figure 3C) and 3677 cells (in the specification at page 40, lines 1-6; page 18, lines 4-7).

Second, Applicants have disclosed the requisite starting materials and the use thereof needed to practice the full scope of the invention e.g., Applicants have disclosed the type of cells that can be used as starting materials for the present invention (in the specification at page 16, lines 15-31; page 24, lines 21-29). In detail, the specification discloses that "any cell can be utilized, since the resulting location of fluorescence can be visualized as either in the cytoplasm or in the nucleus" (in the specification at page 15, lines 19-20). Further the specification describes that "additionally, for such detection

events, cells having increased copy number of the binding site, in an array, can be used" (in the specification at page 15, lines 20-21). Therefore, the present specification clearly teaches that any mammalian cell having a plurality of steroid response elements in an array such that the element can be directly detected can be used to perform the screening for a ligand that activates the translocation of the steroid receptor to the nucleus. By a plurality of steroid response elements is meant that the number of copies of the elements is greater than one (in the specification at page 17, lines 13-23). It has been known for higher eukaryotic genomes to contain naturally occuring repetitive sequences (in the specification at page 61, lines 26-28). Therefore, any mammalian cell which meets these criteria can by used as a starting material for the present invention.

Additionally, the specification describes the localizations of the elements to be "in sufficiently close physical proximity along a chromosome, either present endogenously or artificially introduced or induced, or in extrachromosomally replicating episomes" (in the specification at page 17, lines 3-5; page 19, lines 1-2). Methods for artificially introducing or inducing a gene into chromosomes are well known in the art (Monaco, A.P. and Larin, Z. (1994) *Trends Biotechnol.* **12**, 280-286) (see Evidence Appendix 2), as are transformation or transfection methods for extrachromosomally replicating episomes. Nevertheless, examples for these methods are given in the specification of the invention (e.g., page 40, lines 8-18). A non-limiting example of an artificially modified cell including a high number of copies of the response elements is the cell line 3134 which is described in the application in detail and which is deposited with American Type Culture Collection as accession number CRL-11998 (ATCC) (e.g., in the specification at page 17, lines 22-24). This cell line is particularly useful for detection of ligands. The same results can be achieved by the use of any desired (naturally occurring or artificially modified) mammal cell having a plurality of steroid response elements in such an array, as stated above.

Third, Applicants have disclosed how to prepare additional cell lines containing receptor response elements. Methods to prepare human and other mammalian cell lines containing receptor response elements are disclosed (in the specification at page 57, lines 19 to page 58, line 3) with a detailed step-by-step protocol (in the specification at page 58, lines 5 to page 63, line 13).

Therefore, in view of the reasons stated above, Applicants have provided adequate written disclosure to encompass the scope of the claims i.e., there is adequate written description for the genus of a mammalian cell that has a plurality of steroid response elements.

Adequate Written Description for Use of Any Mammalian Cell with Multiple Steroid Receptor Response Elements

The specification complies with 35 U.S.C. §112 because it provides adequate description as to how any mammalian cell with multiple steroid receptor elements can be used for the method of the claimed invention for the following reasons.

The requirements of 35 U.S.C. §112, first paragraph, are fulfilled where one skilled in the art could use the invention, given the specification disclosure, without undue experimentation[1]. Undue breadth is analyzed in terms of whether it would have involved undue experimentation to achieve the claimed invention. The determination of what constitutes undue experimentation in a given case, requires the application of a standard of reasonableness, having due regard for the nature of the invention and the state of the art[2].

[1] *In Re Eynde,* 480 F.2d 1364, 178 U.S.P.Q. 470 (CCPA 1970) (see Evidence Appendix 3)
[2] *Ex parte Forman, et al.*, 230 U.S.P.Q. 546, 547 (BPAI 1986) (see Evidence Appendix 4)

The test is not merely quantitative, since a considerable amount of experimentation is permissible, if it is merely routine, or if the specification in question provides a reasonable amount of guidance with respect to the direction in which the experimentation should proceed to enable the determination of how to practice a desired embodiment of the invention claimed[3].

In *Ex parte Forman*, the Board set forth the following criteria for undue experimentation:

> The question of undue breadth is analyzed in the view of:
>
> (1) the quantity of experimentation necessary,
>
> (2) the amount of direction or guidance presented,
>
> (3) the presence or absence of working examples,
>
> (4) the nature of the invention,
>
> (5) the state of the prior art,
>
> (6) the relative skill of those in that art, and
>
> (7) the unpredictability of the art[4].

The unpredictability of the art is only one factor that must be evaluated and weighed with the other factors.

The specification fully enable the use of <u>any</u> mammalian cell with multiple steroid receptor response elements in the claimed methods, namely, how to screen for a ligand that activates the translocation of a steroid receptor to the nucleus in a mammalian cell, because of the following reasons.

1. The steroid hormone signal transduction pathway is conserved in all eukaryotic cells, including mammalian cells. Translocation or movement of a steroid receptor from the cytoplasm to the cell nucleus is a necessary step for steroid

[3] *Ex parte Forman, et al.*, 230 U.S.P.Q. 546, 547 (BPAI 1986).

receptors to modulate steroid hormone-responsive gene expression by steroid hormones. The Applicants have demonstrated translocation of the steroid receptor to the nucleus of a mammalian cell (in the specification at page 47, line 1 to page 48, line 14) as part of a screening assay to find ligands that activate the translocation of a steroid receptor to the nucleus in a mammalian cell. In one embodiment, the ligand is dexamethasone and a dose response curve for dexamethasone is provided (in the specification at page 47, lines 1-8). In another embodiment, the ligand is RU486 (in the specification at page 47, lines 26-27; page 48, lines 2-4). In yet another embodiment, the ligand is progesterone (in the specification at page 48, lines 5-7). When a non-ligand for a particular receptor is used to treat a fluorescent receptor, no translocation is observed, demonstrating importance of activating the steroid receptor for translocation to occur from cytoplasm to the nucleus (in the specification at page 48, lines 7-8).

2. The specification provides adequate written description of examples of steroid response elements in an array. For example, the steroid nuclear receptor can be ER, AR, GR, PR, and/or MR (in the specification at page 12, line 19; page 13, line 11; page 18, lines 1-4; page 56, lines 9-17). With regard to ER, Applicants teach the sequence of the steroid response element in the specification at page 59, lines 4-5 and lines 15-17 and methods for making it in the specification at page 59, lines 4-21. With regard to GR, Applicants teach the sequence of the steroid response element in the specification at page 59, lines 3-4 and lines 10-14 and methods for making it in the specification at page 59, lines 3-21. With regard to AR, PR, and MR, the consensus sequence of each of these steroid response elements are the same, and each receptor can bind to the steroid hormone-responsive elements that are recognized by GR.

---

[4] *Forman* at page 547, supra.

Hence, no undue experimentation would be required to practice the claimed invention. Accordingly, Applicants respectfully request that the Office withdraw the rejection.

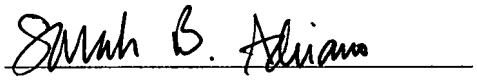
CONCLUSION

Applicants submit that the standards for written description under 35 U.S.C. §112, first paragraph, have been met in view of Applicants' teaching in the specification.

Applicants respectfully request withdrawal of the 35 U.S.C. §112 rejection and allowance of Claims 16 and 17.

No fees are deemed necessary in connection with the filing of this Amended Appeal Brief. If any further fees are necessary, the Patent Office is authorized to charge any additional fees to Deposit Account No. 50-0306.

Respectfully submitted,



Sarah B. Adriano
Registration No. 34,470
Teresa Liang, Ph.D.
Registration No. 51,946
Mandel & Adriano
55 S. Lake Avenue, Suite 710
Pasadena, California 91101
(626) 395-7801
Customer No. 26,941

J. CLAIMS APPENDIX

1 to 15. (Cancelled)

16. (Under Appeal) A method of screening for a ligand that activates the translocation of a steroid receptor to the nucleus in a mammalian cell comprising:

a. contacting a mammalian cell having a nucleus with the ligand, wherein the cell has a plurality of steroid receptor response elements, wherein the steroid receptor response elements comprise a plurality of AGAACA (SEQ ID NO:4) or AGGTCA (SEQ ID NO:5), in an array such that the response element can be directly detected when bound by fluorescently labeled steroid receptor; and

b. detecting the location of fluorescence within the cell,

a change in the relative fluorescence of the nucleus to the cytoplasm so as to increase the fluorescence of the nucleus indicating a ligand that activates the translocation of a steroid receptor to the nucleus in a mammalian cell.

17. (Under Appeal) The method of claim 16, wherein the fluorescently labeled steroid receptor is fluorescently labeled with a green fluorescent protein.

18. (Allowed) A method of screening for a ligand that activates the translocation of a steroid receptor to the nucleus in a mammalian cell comprising:

a. contacting a mammalian cell having a nucleus with the ligand, wherein the cell has a plurality of steroid receptor response elements, wherein the steroid receptor response elements comprise a plurality of AGAACA (SEQ ID NO:4) or AGGTCA (SEQ ID NO:5), in an array such that the response element can be directly detected when bound by fluorescently labeled steroid receptor; and

b. detecting the location of fluorescence within the cell,

a change in the relative fluorescence of the nucleus to the cytoplasm so as to increase the fluorescence of the nucleus indicating a ligand that activates the translocation of a steroid receptor to the nucleus in a mammalian cell,

wherein the mammalian cell is a cell of the cell line designated 3134 deposited with American Type Culture Collection under accession number CRL-11998 (ATCC).

19 to 58. (Cancelled)

K. EVIDENCE APPENDIX

1. Application U.S. Serial Number 10/001,486 – subject application with Claims 16 and 17 under appeal.
2. Monaco, A.P. and Larin, Z. (1994) *Trends Biotechnol.* **12**, 280-286 – cited in originally filed application U.S. Serial Number 10/001,486 at page 18 line 28 to page 19 line 2.
3. *In Re Eynde,* 480 F.2d 1364, 178 U.S.P.Q. 470 (CCPA 1970) – cited in the Office Action response dated May 2, 2006 at page 5 (footnote 3).
4. *Ex parte Forman, et al.*, 230 U.S.P.Q. 546, 547 (BPAI 1986) – cited in the Office Action response dated May 2, 2006 at pages 5-6 (footnotes 4-6).

EVIDENCE APPENDIX 1

U.S. Serial No. 10/001,486
Originally filed specification

METHOD AND COMPOSITIONS FOR MONITORING DNA BINDING MOLECULES IN LIVING CELLS

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates to the field of binding of molecules such as transcription factors to regions of nucleic acids, steroid hormone usage, steroid receptors and their corresponding response elements. Reagents are provided to allow methods involving direct detection of binding of a molecule, determining response element targeting by activated steroid receptors, screening for steroid agonists and antagonists, and monitoring levels of steroid agonists and antagonists in biological samples.

Background Art

Steroid receptors are hormone-dependent activators of gene expression. Steroid receptors mediate the action of steroid hormones (*e.g.*, glucocorticoids, estrogens, progestins, testosterone, mineralocorticoids and 1,25-dihydroxycholecalciferol) in human tissues. After activation with the cognate ligand, receptors bind to chromatin in the nucleus and modulate the activity of target cellular genes. The binding of receptors to these target sequences is a key step in steroid function. Currently, this interaction can only be detected by indirect methods, such as reporter assays that detect the result of transcriptional activation coupled with transfection methods that introduce DNA sequences with receptor binding sites.

It is generally accepted that the unliganded glucocorticoid receptor (GR) resides in the cytoplasm, and that hormone activation leads both to nuclear accumulation and gene activation. (Gasc, J. -M. & Baulieu, E. E. (1987) in *Steroid Hormone Receptors: Their Intracellular Localisation*, ed. Clark, C. R. (Ellis Horwood Ltd., Chichester, England), pp. 233-250; Beato, M. (1989) *Cell* **56**, 335-344; Carson-Jurica, M. A., Schrader, W. T. & O'Malley, B. W. (1990) *Endocr. Rev.* **11**, 201-220; Gronemeyer, H. (1993) in *Steroid Hormone Action*, ed. Parker, M. G. (Oxford University Press, New York), pp. 94-117; Tsai, M. J. & O'Malley, B. W. (1994) *Annu. Rev. Biochem.* **63**,

451-486; Akner, G., Wikstrom, A. C. & Gustafsson, J. A. (1995) *J. Steroid Biochem. Mol. Biol.* **52**, 1-16), and references therein. However, the mechanisms involved in nuclear translocation and targeting of steroid receptors to regulatory sites in chromatin have been poorly understood. It has previously been difficult to discriminate between the ability of a given receptor mutant, or a given receptor/ligand combination, to participate in the separate processes of receptor activation, nuclear translocation, sequence-specific binding, and promoter activation.

Proteins have previously been labeled with fluorescent tags to detect their localization and sometimes their conformational changes both *in vitro* and in intact cells. Such labeling is essential both for immunofluorescence and for fluorescence analog cytochemistry, in which the biochemistry and trafficking of proteins are monitored after microinjection into living cells (Wang, Y.L. & Taylor, D.L., eds. (1989) *Methods Cell Biol.* **29**). Traditionally, fluorescence labeling is done by purifying proteins and then covalently conjugating them to reactive derivatives of organic fluorophores. The stoichiometry and locations of dye attachment are often difficult to control, and careful repurification of the proteins is usually necessary. If the proteins are to be used inside living cells, a final challenging step is to get them across the plasma membrane via micropipet techniques or various methods of reversible permeabilization. Furthermore, in previous hormone studies broken cell preparations or antibody tags in fixed cell preparations were used, both techniques that cause enormous disruption of cell structures.

The green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* is a molecule whose natural function seems to be to convert the blue chemiluminescence of the Ca^{2+}-sensitive photoprotein aequorin into green emission (Ward,W.W. (1979) in *Photochemical and Photobiological Reviews*, ed. Smith, K.C. (Plenum, New York), 4:1-57). GFP's absorption bands in the blue (maximally at a wave length of 395nm with weaker absorbance at 470nm) and emission peak in the green (at 509mn) do not arise from a distinct cofactor but rather from an internal *p*-hydroxybenzylideneimidazolidinone chromophore generated by cyclization and

oxidation of a serine-tyrosine-glycine sequence at residues 56-67 (Cody, C.W., Prasher, D.C., Westler, W.M., Prendergast, F.G. & Ward, W.W. (1993) *Biochemistry* **32**, 1212-1218). The gene for GFP was cloned (Prasher, D.C., Eckenrode, V.K., Ward, W.W., Prendergast, F.G. & Cormier, M.J. (1992) *Gene* **111**, 229-233), and the encoded protein consists of 238 amino acid residues (molecular weight 27kD). Heterologous expression of the gene has been done in *Escherichia coli* (Heim, R., Prasher, D.C. and Tsien, R.Y. (1994) *Proc. Natl. Acad. Sci.U.S.A.* **91**,12501-12504) ; Inouye, S. & Tsuji, F.I. (1994) *FEBS Lett.* **341**, 277-280), *Caenorhabditis elegans* (Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W. & Prasher, D.C. (1994) *Science* **263**, 802-805), and *Drosophila melanogaster* (Yeh, E., Gustafson, K. & Boulianne, G.L. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7035-7040; Tannahill, D., Bray, S. & Harris, W.A. (1995) *Dev. Biol.* **168**, 694-697 and plants (Hu,W. & Cheng, C.L. (1995) *FEBS Lett.* **369**, 331-334; Baulcombe, D.C., Chapman, S. & Santa Cruz, S. (1995) *Plant J.* 7, 1045-1053). Recently, chimeric genes encoding N- and C-terminal fusions of the Drosophila exuperantia (exu) gene product, Exu (Wang, S. and Hazelrigg, T. (1994) *Nature* **369**, 400-403), actin Act88F gene (Barthmaier, P. and Fyrberg, E. (1995) *Dev. Biol.* **169**, 770-774), and a nuclear localization signal (Davis, I., Girdham, C.H. & O'Farrell, P.H. (1995) *Dev. Biol.* **170**, 726-729); of the yeast microtubule and spindle pole associated dis1 gene product (Nabeshima, K., Kurooka, H., Takeuchi, M., Kinoshita, K., Nakaseko, Y., & Yanagida, M. (1995) *Genes Dev.* **9**, 1572-1585) and an RNA binding protein Npl3p (Corbett, A.H., Koepp, D.M., Schlenstedt, G., Lee, M.S. Hopper, A.K. & Silver, P.A. (1995) *J. Cell Biol.* **130**, 1017-1026); and of a mammalian ion channel protein, NMDAR1 (Marshall, J., Molloy, R., Moss, G.W., Howe, J.R. & Hughes, T.E. (1995) *Neuron* **14**, 211-215), microtubule-associated protein, MAP4 (Olson, K.R., McIntosh, J.R. & Olmsted, J.B. (1995) *J. Cell Biol.* **130**, 639-650), and a secretory protein, chromogranin B (Kaether, C. & Gerdes, H.H. (1995) *FEBS Lett.* **369**, 267-271) have been constructed fused to GFP. However, none of these chimeric proteins have been to transcription factors or co-factors and no suggestions have been made as to the usefulness of such a fusion to study physiologically relevant interaction on an amplified

DNA target. Furthermore, none of these reports indicated a successful use of GFP in mammalian cells.

Many human diseases result from aberrant steroid function, and many disease states, *i.e.*, inflammation, are treated with glucocorticoid and other steroid derivatives. A large number of drugs have been developed whose function is based on the ability to interact with and activate steroid receptors. The identification and characterization of these compounds is a laborious, time-consuming and expensive process involving years of work. Even with a large investment of resources, the true behavior of these compounds in living cells is not understood.

The present invention allows observation for the first time of *in vivo* target sites within a higher eukaryotic nucleus for trans-regulatory molecules, such as transcription factors, *e.g.*, glucocorticoid receptor (GR). The visualization of physiologically relevant *in vivo* target sites for any transcription factor to date has not previously been accomplished. The present invention provides a poweful method for identification of any single target site in a higher eukaryotic genome, comprising roughly 60,000-80,000 genes (Bird, A.P. (1995) *Trends Genet.* 11:94-100), using a singly fluorescently-labelled regulatory factor, which has not been considered previously. Discriminating direct versus indirect interaction between a regulatory molecule and its putative regulatory site is critical for the development of highly specific drugs directed against trans-regulatory factors. Traditionally, the methodology for showing potentially direct interactions involves nuclease or chemical protection experiments and transient co-transfection experiments of the putative regulator and its regulated site. While this approach indicates potential direct interaction, it does not necessarily imply direct interaction. Alternatively, the approach of making compensatory mutations between the regulatory sequences as well as the DNA binding specificity has been used in an attempt to demonstrate direct regulatory interaction (Schier, A.F. and Gehring, W.J. (1992) *Nature* 356:804-807), an extension of the principles of second site suppression in genetics to molecular biology. However, such an approach makes enormous assumptions of our understanding of sequence-specific recognition by sequence-specific DNA binding

proteins *in vivo*, which certainly would not be valid for many systems, since many profound developmental events are governed by exquisite interactions to fine tune the system regarding, for example, concentration gradients of trans-regulatory factors. The present invention allows a simple and straight-forward manner in which direct interaction between a sequence-specific DNA binding protein or its co-factor and its putative regulatory site in the *in vivo* genomic context can be addressed. With this simple inventive methodology, novel classes of drugs directed not only against members of the steroid-ligand-dependent transcription factors but to new classes of drugs that target other transcription factors or their co-factors can be screened.

Additionally, the present invention provides the first opportunity to observe and monitor gene targeting specifically of steroid receptors in living cells wherein binding of the steroid receptor to its response element target can be observed distinctly from translocation of steroid receptor. The invention therefore provides for many relevant analyses, such as real-time determination of steroid activity in subjects as well as screening of compounds for response element binding/targeting capabilities as distinct from translocation capabilities. Such methods have implications in many diseases associated with steroid hormones, such as endocrine disorders, rheumatic disorders, collagen diseases, dermatological diseases, allergic states, ophthalmic diseases, respiratory disease, hematologic disorders, neoplastic disease, edematous states, gastrointestinal diseases and neurological conditions, and in other uses such as prevention of rejection of transplanted tissues.

SUMMARY OF THE INVENTION

The present invention provides a mammalian cell having a plurality of steroid receptor response elements in an array such that the response element can be directly detected when bound by fluorescently labeled steroid receptor.

The present invention further provides a chimeric protein comprising a fluorescent protein fused to a transcription factor. The present invention also provides a chimeric protein comprising a fluorescent protein fused to a steroid receptor.

The instant invention provides an isolated nucleic acid encoding a chimeric protein comprising a fluorescent protein fused to a transcription factor and an isolated nucleic acid encoding a chimeric protein comprising a fluorescent protein fused to a steroid receptor.

The instant invention also provides a cell containing a nucleic acid encoding a chimeric protein comprising a fluorescent protein fused to a transcription factor and a cell containing a nucleic acid encoding a chimeric protein comprising a fluorescent protein fused to a steroid receptor.

The instant invention provides a method of screening for a compound that binds to a selected nucleic acid comprising:

a. contacting compound fluorescently labeled by a fluorescent protein with a cell having a plurality of copies of the nucleic acid in an array such that the nucleic acid can be directly detected when bound by fluorescently labeled compound; and

b. directly detecting the location of fluorescence within the cell, fluorescence aggregated at the site of the nucleic acid array indicating a compound that binds to the selected nucleic acid.

The present invention also provides a method of characterizing a ligand's effect on cellular localization of a compound to which the ligand binds in a cell comprising:

a. contacting the ligand with a cell having the compound fluorescently labeled by a fluorescent protein and

b. directly detecting the location of fluorescence within the cell,

the location of fluorescence in the cell indicating the localization effect of the ligand on the compound.

Additionally provided is a method of determining a binding site for a DNA-binding protein comprising:

a. contacting the DNA-binding protein fluorescently labeled by a fluorescent protein with a cell having a plurality of copies of a nucleic acid having a putative binding site in an array such that the putative binding site can be directly visualized when bound by the fluorescently labeled DNA-binding protein, and

b. directly detecting the location of fluorescence within the cell,

the presence of fluorescence aggregated at the putative binding site indicating a binding site to which the DNA-binding protein binds.

The present invention also provides a method of screening for a ligand that activates gene targeting of a steroid receptor in the nucleus of a mammalian cell comprising:

a. contacting the ligand with a mammalian cell having a plurality of steroid receptor response elements in an array such that the response element can be directly detected when bound by fluorescently labeled steroid receptor and the cell further comprising a nucleic acid encoding a chimeric protein wherein a fluorescent protein is fused to the steroid receptor; and

b. directly detecting the location of fluorescence within the cell,

fluorescence aggregated at the site of the steroid receptor response element array in the nucleus indicating a ligand that activates the gene targeting of a steroid receptor in the nucleus of a mammalian cell.

The present invention provides a method of screening for a ligand that activates the translocation of a steroid receptor to the nucleus in a mammalian cell comprising:

a. contacting the ligand with a mammalian cell having a plurality of steroid receptor response elements in an array such that the response element can be directly detected when bound by fluorescently labeled steroid receptor and the cell further comprising a nucleic acid encoding a chimeric protein wherein a fluorescent protein is fused to the steroid receptor; and

b. directly detecting the location of fluorescence within the cell,

the location of fluorescence aggregated in the nucleus indicating a ligand that activates the translocation of a steroid receptor to the nucleus in a mammalian cell.

The instant invention provides a method of detecting in a biological sample the presence of an agonist of a steroid receptor comprising:

a. contacting the sample with a mammalian cell having a plurality of steroid receptor response elements in an array such that the response element can be directly detected when bound by fluorescently labeled steroid receptor and the cell further comprising a nucleic acid encoding a chimeric protein wherein a fluorescent protein is fused to the steroid receptor; and

b. directly detecting the location of fluorescence within the cell,

the location of fluorescence aggregated at the site of the steroid receptor response element array in the nucleus indicating the presence of an agonist of the steroid receptor in the sample.

The present invention also provides a method of detecting in a biological sample the presence of an antagonist of a steroid receptor comprising:

a. contacting the sample and an agonist of the steroid receptor with a mammalian cell having a plurality of steroid receptor response elements in an array such that the response element can be directly detected when bound by fluorescently labeled steroid receptor and the cell further comprising a nucleic acid encoding a chimeric protein wherein a fluorescent protein is fused to the steroid receptor; and

b. directly detecting the location of fluorescence within the cell,

the absence of fluorescence substantially aggregated at the site of the steroid receptor response element array in the nucleus indicating the presence of an antagonist of the steroid receptor in the sample.

The present invention provides a method of monitoring the level of an agonist of a steroid receptor in a subject comprising:

a. periodically obtaining a biological sample from the subject,

b. contacting the sample with a mammalian cell having a plurality of steroid receptor response elements in an array such that the response element can be directly detected when bound by fluorescently labeled steroid receptor and the cell further comprising a nucleic acid encoding a chimeric protein wherein a fluorescent protein is fused to the steroid receptor; and

c. directly detecting the location of fluorescence within the cell,

a decrease in fluorescence aggregated at the site of the steroid receptor response element in the nucleus in a later-obtained sample relative to an earlier-obtained sample indicating a decrease in level of the steroid agonist of the steroid receptor in the sample and an increase in fluorescence aggregated at the site of the steroid receptor response element in the nucleus in a later-obtained sample relative to an earlier-obtained sample indicating an increase in level of the steroid agonist of the steroid receptor in the sample.

The instant invention provides a method of monitoring the balance between levels of an agonist of a steroid receptor and an antagonist of the steroid receptor in a subject comprising:

a. periodically obtaining a biological sample from the subject,

b. contacting the sample with a mammalian cell having a plurality of steroid receptor response elements in an array such that the response element can be directly detected when bound by fluorescently labeled steroid receptor and the cell further comprising a nucleic acid encoding a chimeric protein wherein a fluorescent protein is fused to the steroid receptor; and

c. directly detecting the location of fluorescence within the cell,

an increase in fluorescence aggregated at the site of the steroid receptor response element in the nucleus in a later-obtained sample relative to an earlier-obtained sample indicating an increase in level of the steroid agonist relative to level of the steroid antagonist in the sample, and a decrease in fluorescence aggregated at the site of the steroid receptor response element in the nucleus in a later-obtained sample relative to an earlier-obtained sample indicating an increase in level of the steroid antagonist of the steroid receptor relative to level of the steroid agonist in the sample.

The instant invention also provides a method of determining an effective dosage of a steroid receptor agonist in a subject comprising:

a. transferring into a set of cells from the patient a nucleic acid encoding a chimeric protein comprising a fluorescent protein fused to a steroid receptor;

b. contacting the cells in the set with one of a selected range of dosages of the steroid agonist; and

c. directly detecting location of fluorescence in the set of cells, a dosage capable of locating fluorescence substantially in the nucleus indicating an effective dosage of steroid receptor agonist.

The present invention provides a method of determining an effective dosage of a steroid receptor agonist to maintain steroid receptor activation for a selected period of time in a subject comprising:

a. administering to the subject a dosage of the steroid receptor agonist,

b. periodically obtaining a biological sample from the subject,

c. contacting the sample with a mammalian cell having a plurality of steroid receptor response elements in an array such that the response element can be directly detected when bound by fluorescently labeled steroid receptor and the cell further comprising a nucleic acid encoding a chimeric protein wherein a fluorescent protein is fused to the steroid receptor; and

d. directly detecting the location of fluorescence within the cell,

a dosage that maintains the location of fluorescence at the site of the steroid receptor response element array in the nucleus for the selected period of time indicating an effective dosage.

The present invention also provides a method of determining an effective dosage of a steroid receptor antagonist to abrogate agonist activity for a selected period of time in a subject comprising:

a. administering to the subject a dosage of the steroid receptor agonist,

b. periodically obtaining a biological sample from the subject;

c. contacting the sample with a mammalian cell having a plurality of steroid receptor response elements in an array such that the response element can be directly detected when bound by fluorescently labeled steroid receptor and the cell further comprising a nucleic acid encoding a chimeric protein wherein a fluorescent protein is fused to the steroid receptor; and

d. directly detecting the location of fluorescence within the cell,

a dosage that prevents the location of fluorescence at the site of the steroid receptor response element array in the nucleus for the selected period of time indicating an effective dosage.

The present invention also provides a method of detecting a defect in a response pathway of a steroid receptor in a subject comprising transferring into a cell from the subject a nucleic acid functionally encoding a chimeric protein comprising a fluorescent

protein fused to the steroid receptor and detecting the location of fluorescence within the cell as compared to the location of fluorescence within a normal, control cell transfected with the nucleic acid, a difference in location of fluorescence within the cell of the subject as compared to location of fluorescence within the normal, control cell indicating a defect in the response pathway of the steroid receptor.

The instant invention provides a method of determining whether a defect in a response pathway of a steroid receptor in a subject is in translocation of the steroid receptor to a cell nucleus, comprising transferring into a cell from the subject having the defect a nucleic acid functionally encoding a chimeric protein comprising a fluorescent protein fused to the steroid receptor and detecting the location of fluorescence within the cell, the location of fluorescence substantially in the cytoplasm of the cell indicating the defect is in translocation of the steroid receptor to the nucleus.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a diagram of the integrated, tandem array of bovine papilloma virus (BPV) 69% transforming fragment, mouse mammary tumor virus (MMTV) long terminal repeat (LTR) containing glucocorticoid receptor (GR) binding sites, and Harvey murine sarcoma virus v-ras sequences (Ha-v-ras).

Figure 2 shows, in panels 11-17, the appearance of GFP-GR when the fluorescent tag is excited by 489 nm laser light, and the 511 nm emission examined by confocal microscopy, using a standard fluorescein filter set. Individual sections are depicted for a representative 3134 cell nucleus. As one focuses on 0.18 micrometer sections throughout the nucleus, a continuous fiber of intense light emission is detected over 5-7 sections, which corresponds to GFP-GR binding to the continuous BVP/MMTV-LTR/ras array. The array is seen in sections 12-16. Below the panels is provided a schematic presentation of the appearance of GFP-GR when the fluorescent protein is excited.

Figure 3 shows the construction of GFP-GR. (A) Plasmid pCI-nGFP-C656G contains the green fluorescent protein fused to the C656G mutant glucocorticoid

receptor. (**B**) Dexamethasone (dex) dependent stimulation of MMTV-pLTRLuc is shown for GFP-GR transfected cells. Solid bars represent the IL2R$^+$ selected population activated with 1 nM dex, and the open bar depicts activation of the endogenous receptor with 100 nM dex. (**C**) Ligand specificity is presented for activation of endogenous MMTV-LTR-CAT sequences present in the 1471.1 cells. Cells were treated for 4 hrs with the indicated ligand, then harvested and levels of CAT activity determined.

Figure 4 shows nuclear localization of GFP-GR in MCF7 cells when the cells are treated with (A) dexamethasone or (B) RU486. The fluorescent tag is excited by 489 nm laser light, and the 511 nm emission examined by confocal microscopy, using a standard fluorescein filter set.

Figure 5 shows nuclear localizationof GFP-ER in (A) MCF7 cells and (B) MDA-MB-231 cells when the fluorescent tag is excited by 489 nm laser light, and the 511 nm emission examined by confocal microscopy, using a standard fluorescein filter set. MCF7 cells are reported to be estrogen receptor positive and hormone dependent. MDA-MB0231 cells are reported to be estrogen receptor negative and hormone independent.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides cell lines such as the murine cell line 3134, that contains a set of mouse mammary tumor virus (MMTV) Harvey murine sarcoma virus (HaMuSV) v-ras sequences organized in a head-to-tail tandem array of approximately 200 copies. Each MMTV promoter sequence in this array contains 4 glucocorticoid receptor (GR) binding sites; the complete array thus contains nearly 1000 GR binding sites. This cell is used to visualize directly the interaction between GR and its binding site in chromatin in living cells. This is accomplished with a fluorescent labeled copy of the GR. This array thus provides the unique opportunity to visualize direct receptor/target interactions. This structure, for which there is no precedent, provides for the first time a reagent to observe the interaction of steroid receptors with their response elements in living cells, and to characterize the effectiveness of medically important steroid ligands in activating gene expression in mammalian cells.

In the present application, we have shown the usefulness of GFP in monitoring the activity of a steroid hormone receptor. The ability to directly observe living cells has allowed us to follow in real time the process of cytoplasm to nuclear translocation, and has revealed for the first time differences in GR intranuclear accumulation pattern dependent on the type of activating ligand. Furthermore, the patterns of GR accumulation are remarkably similar between adjacent cells, suggesting an order in the organization of the interphase nucleus. This ability to observe living cells has also revealed subcellular localization of partially activated estrogen receptor (ER). Thus, the use of GFP has revealed new details about steroid localization and organization of the eukaryotic nucleus.

As used in the specification and in the claims, "a" can mean one or more, depending upon the context in which it is used.

The present invention provides reagents and methods for detecting, by direct visual evaluation, the binding of fluorescently labelled compounds to a nucleic acid. This direct detection can be accomplished by the use of a cell line having a sufficient number of copies of the binding region of the nucleic acid in an array, such as tandem repeats, that allows detection of the array when a fluorescently labeled compound is bound to the binding region by direct detection of the fluorescence localized at the site in the cell nucleus of the nucleic acid array. For example, after binding, cells can immediately, without further treatment of the cells, be placed under a fluorescent microscope and fluorescence directly visualized. Thus compounds can be rapidly analyzed for binding capability in a manner that clearly depicts the binding. Importantly, the binding detected in this assay is biologically relevant. The detected event represents receptor mobilization to a correct genetic target in the living cell. Thus, interactions revealed with this assay are much more reliable as measures of biologically germane receptor activity. Additionally, analyses of levels of compounds or defects in pathways involving the binding of such compounds to their nucleic acid binding site in specific subjects can be performed, as further described below.

Any selected nucleic acids and binding sites, and compounds that may bind thereto, either directly or indirectly, can be analyzed by this method, in a desired cell, as elaborated below. For example, binding of any transcription factor to its activation site on a nucleic acid can be directly determined along with the proteins associated with the factor (e.g., co-activator, co-repressor, adapter, or molecules in a similar category). Additionally, for example, binding of any selected steroid receptor can be directly detected and analyzed for both translocation to the nucleus and binding to the corresponding response element(s) in the nucleus. Thus, compounds typically can be proteins, polypeptides and peptides; however, other compounds can include, for example, Peptide Nucleic Acids (PNAs), antisense nucleic acids and organic molecules (e.g., dexamethasone). Importantly, for any compound, the step of binding to the binding site can visually be distinguished from the step of translocation to the nucleus, or alternatively, exit from the nucleus.

In general, the present invention utilizes fluorescent labeling of the compound by a fluorescent protein, as fluorescent protein is herein described, adding the labeled compound to cells, and directly detecting the location and/or aggregation of fluorescence in the cells. For detection of the translocation of the labeled compound to the nucleus, any cell can be utilized, since the resulting location of fluorescence can be visualized as either in the cytoplasm or in the nucleus. Additionally, for such detection events, cells having increased copy number of the binding site, in any array, can be used. For detection of binding to the target nucleic acid site, the present invention provides cells having a plurality of nucleic acid binding sites in an array such that the nucleic acid binding site can be directly detected when bound by binding compound, such as a ligand, transcription factor, etc., fluorescently labeled by a fluorescent protein, as described herein.

As used herein, direct detection means detection of the fluorescence emitted from the site in the cells when excited by light, ultraviolet or visible, without the need for any additional chemical reactions or treatment of the cells. The fluorescence is directly detected by any device capable of detecting fluorescence, such as a fluorescent

microscope, as visualized by the eye of the operator of the microscope at the time or as recorded from the microscope such as by photography of the field of view or through the use of photosensitive detectors. A fluorescent microscope, such as a confocal laser scanning microscope or an epifluorescent microscope, can be used, as is known in the art. There is no requirement that cells be, *e.g.*, fixed or stained or contacted by any additional reagents, in order to detect the binding. Thus living cells can be assayed, and results obtained, immediately after binding. Therefore, for example, subjects can be advised immediately of results of analyses as described below. Furthermore, it is anticipated that screening of both nuclear localization of fluorescence (translocation) and focal localization of fluorescence on a target array will be adapted to high volume computerized image analysis. That is, the analysis of large numbers of samples will be automated for either the repetitive examination of clinical samples or the large-scale screening of compounds in the research environment.

The cell can be derived from any desired mammal, such as, for example, human, monkey, mouse, hamster and rat. The nucleic acid can be amplified in an appropriate array by any of several means, as known to those skilled in the art. Generally, a selected nucleic acid binding site or collection of sites, for example as found within the context of a transcriptional regulatory region, i.e., promotors, enhancers, silencers, etc., can be amplified in an array detectable by the present means, for example, by gene amplification of the nucleic acid binding site (e.g., the steroid receptor response element or the transcription factor binding site) by linking it to a gene readily amplified in a tandem array, for example, dihydrofolate reductase, or by multimerization of the nucleic acid binding site or sites by synthetic DNA synthesis and/or enzymatic synthesis, for example, through the use of ligase and polymerases, and introducing the amplified element into selected mammalian cells. Such methods are further elaborated in the examples provided below.

The cells used herein have an array of the nucleic acid having binding sites under analysis such that the nucleic acid can be directly detected when bound by a fluorescently labeled compound. Such array as contemplated herein has sufficient copies

of the nucleic acid and in such an arrangement that the fluorescently labeled binding compound, when bound to the site can be directly detected and readily identified. Therefore, the array includes arrangement of the copies in sufficiently close physical proximity along a chromosome, either present endogenously or artificially introduced or induced, or in extrachromosomally replicating episomes, to allow localization of fluorescence at a discrete, detectable site in the nucleus, as seen under standard magnification for cells and nuclei. Such array as contemplated herein allows detection in the context of chromatin, as exists in the interphase nucleus. An example of such an array is a series of direct tandem repeats of the nucleic acid. An example of a tandem array of direct repeating units is depicted in FIG. 1.

By a "plurality" of any herein described nucleic acid having a binding site is meant that the number of copies of the nucleic acid having the binding site (e.g., the steroid receptor response element or the transcription factor binding site) is greater than one. Preferably, the cells have more than about five copies, more preferably more than about ten copies, more preferably more than about twenty, and more preferably more than about forty copies. For example, cell line 3134, described herein, has about two hundred copies of the MMTV LTR-Ha-v-ras- gene, each of which has four copies of the binding site for glucocorticoid receptor and with each site accommodating two glucocorticoid receptor molecules. Any number which allows detection of the site upon binding of the fluorescently labeled binding compound is contemplated. Thus, an example of a cell of the present invention is a cell of the cell line 3134 deposited with American Type Culture Collection as accession number CRL-11998 (ATTC).

Specifically, the present invention provides a cell having a plurality of steroid receptor response elements in an array such that the response element can be directly detected when bound by fluorescently labeled steroid receptor. A response element, as used herein, includes any nucleic acid to which a steroid receptor directly binds, but also includes the steroid receptor associated, either directly or indirectly, factors that are recruited to the vicinity of the element (e.g., nuclear factor 1 (NF1), octomer transcription factor 1 (OTF1), steroid receptor coactivator 1 (SRC1), etc.). Steroid

receptors, and corresponding response elements to which they bind, can include any steroid receptor, for example, glucocorticoid receptor, estrogen receptor (ER), progesterone receptor, androgen receptor, mineralocorticoid receptor, vitamin D receptor. Examples of steroid receptor response elements include those contained in the mouse mammary tumor virus (MMTV) long terminal repeat (LTR) (which has binding sites for glucocorticoid receptor, mineralocorticoid receptor, progesterone receptor, and androgen receptor), and those contained in vitellogenin and osteocalcin genes (which have binding sites for estrogen and vitamin D receptors). Thus, response elements in such array in a cell can include other transcriptional regulatory elements contained within the mouse mammary tumor virus long terminal repeat and bovine pappilloma virus 69% transforming DNA. Many steroid receptors and steroid response elements, as exemplified above, are known to the skilled artisan; however, any steroid receptor and its response element is contemplated herein. In addition to steroid receptors, there are other ligand-dependent receptors (such as thyroid hormone receptor, retinoic acid receptor, retinoid X receptor, TCCD (dioxin) receptor, fatty acid activatable receptors, etc.) and stimulus-dependent receptors (such as peroxisome proliferator activated receptor, growth factor-dependent receptors (e.g., epidermal growth factor, nerve growth factor, etc.)), and factors (such as CREB, NFAT, NFkB/IkB, etc.), and other receptors whose ligand remains to be defined (such as mammalian homologs of the Drosophila tailless, knirps, sevenup, FTZF1 genes, etc.). Many of these receptors or factors can be found listed in the book [Parker, M.G. (1993) *Steroid Hormone Action* (Oxford University Press, New York, pp. 210)], in a recent review article [Tsai, M.J. & O'Malley, B.W. (1994) *Annu. Rev. Biochem.* **63**, 451-486], and in the GenBank database, which will contain additional receptors as well as the complete nucleotide sequences of the genes and cDNAs. In addition, the cell line offers a number of tandemly repeated regulatory sites for sequence-specific transcription factors (such as activating protein 2 (AP2), OTF1, NF1/CTF, etc.) as well as general transcription factors (such as TFIID, initiator protein, etc.). The steroid receptor response elements (or other transcriptional regulatory elements) used in the present invention in arrays detectable as described herein can be integrated into the genome of the cell, maintained

in the cell on artificial mammalian chromosomes (Monaco, A.P. and Larin, Z. (1994) *Trends Biotechnol.* **12**, 280-286) or can be carried on episomal elements.

The cell can further comprise a nucleic acid encoding a chimeric protein wherein a fluorescent protein is fused to the steroid receptor. Steroid receptors as described above are contemplated for use in a chimeric protein. Typically, the steroid receptor in the chimeric protein encoded by the nucleic acid in a selected cell is one that binds to the response element in the selected cell in the herein described array. Throughout this application by "a fluorescent protein" is meant a protein that fluoresces in cells without adding exogenous cofactors. That is, it is a protein that can be expressed in cells and detected in these cells simply by exciting the protein with light and visualizing the resultant fluorescence. An example of such a fluorescent protein is the green fluorescent protein (GFP) originally isolated from the jellyfish *Aequorea victoria*. Another example of a fluorescent protein as defined herein is the green fluorescent protein originally isolated from *Renilla reniforms*, which demonstrated a single absorption peak at 498 nm and an emission peak at 509 nm. (Cubitt, *et el.* (1995) TIBS 20: 448-455). By fluorescent protein is also contemplated that modifications may be made to a fluorescent protein, as long as the resulting protein fluoresces when expressed in cells. Modifications can be developed based upon the chemistry of chromophore formation. (Cubitt, *et al.* TIBS 20: 448-455). Generally, however, one may prefer to leave the glycine which participates in forming the chromophore, in part by cyclization of a Ser-Tyr-Gly moiety (Gly 67 in *Aequorea* GFP in the Ser65-Tyr66-Gly67 moiety), intact. An example of a useful substitution that modifies the absorption spectra is the substitution in *Aequorea* green fluorescent protein for serine at amino acid 65 by, for example, threonine, cysteine, leucine, valine, or alanine, that allows the excitation of the chromophore at a lower energy (longer wavelength) than the naturally occurring protein thereby greatly decreasing the destruction of the chromophore as occurs when it is excited at a higher energy. Such mutation at amino acid 65 also increases brightness and rate of oxidation as compared to wild-type *Aequorea* GFP when each is excited at its longest wavelength peak. In addition, other spectral variants of GFP, such as improved blue variants of GFP have been developed (*e.g.*, pCI-nGL2-C656G; pCI-nGL3-C656G;

pCI-nGL4-C656G; pCI-nGL5-C656G; pCI-nGL7-C656G; pCI-nGL9-C656G; pCI-nGL10-C656G; pCI-nGL11-C656G). Variants emitting longer wavelengths (*e.g.*, red variants) can also be developed by introducing other mutations into the GFP DNA. Additionally, the codon usage of any GFP- coding sequence can be modified to human codons, according to known methods. For example, pGreenLantern-1 (LifeTechnologies, Inc., Gaithersburg, MD, catalog number 10642-015)is a commercially available S65T variant GFP cDNA with mammalian codon usage. Other commerically available humanized GFP-cDNAs are: pEGFP-N (catalog numbers: 6086-1; 6085-1; 6081-1 from Clontech), pEGFP-C (catalog numbers: 6084-1; 6083-1; 6082-1 from Clontech), and pHGFP-S65T (catalog number 6088-1 from Clontech). Additional useful modifications of any fluorescent protein can include other modifications that speed up the rate of the oxidation step of chromophore formation, that increase brightness at longer wavelengths, and that reduce pohotoisomerization and/or photobleaching. Furthermore, in general it is preferable that GFP not be truncated by more than about one amino acid from the amino terminus and about 10-15 amino acids from the carboxyl terminus. Detection of additional fluorescent proteins can readily be performed by standard approaches such as searching for proteins having some homology to GFP in nucleic acid libraries from organisms that demonstrate fluorescence by nucleic acid hybridization and by searching for homologous nucleic acids and proteins in other organisms in databanks of nucleic acid and protein sequences and testing the encoded proteins for fluorescence. It is possible that forced protein evolution of the currently existing GFP can be achieved by randomizing the entire GFP coding region so as to make every single possible change at every single amino acid coding region as well as pairs and further combination of changes. Desirable changes yielding better chromophore or different excitation/emission spectra can be characterized by fluorescence spectroscopy or flow cytometry upon translation of the coding sequences into proteins. An example of one general approach would be to take advantage of the phage display system for expression of the chromophore on the surface of a bacteriophage using a modification of what is currently being done for antibodies (e.g. Pharmacia Biotech, Inc.'s Recombinant Phage Antibody System). Another general approach would be to adapt a protocol similar to that used to select novel enzymatic

activities displayed by RNAs (Bartel, D.P. and Szostak, J.W. (1993) *Science* **261**, 1411-1418) for selection of novel fluorescent proteins.

Therefore, also provided herein are chimeric proteins comprising a fluorescent protein fused to a transcription factor, and nucleic acids encoding such proteins. One example of a transcription factor that can be used herein, described in further detail above, is a steroid receptor; however, numerous other transcription factors can be utilized. For example, basal transcription factors (e.g. TFIID, etc.), and sequence specific DNA binding transcription factors (e.g., AP1, AP2, SP1, NF1, etc). Additional transcription factors are listed in, for example, computer databases such as that maintained by the National Center for Biotechnology Information (NCBI, Bethesda, MD) accessible through the BLAST program (see item 19 (TFD) for transcription factors; item 20 for eukaryotic promoter sequences). Additionally, as used in the claims, "transcription factors" include transcription adaptor molecules or cofactors, which localization within the cell can be monitored also by this method. Transcription adaptor molecules or cofactors are those molecules that interact with transcription factors to effect their function (*i.e.*, their activation or repression functions). For example, SRC1, steroid receptor coactivator 1, is a cofactor of steroid receptors.

The chimeric protein can include a linking peptide sequence between the fluorescent protein and the steroid receptor. For example, a sequence of the amino acids glycine and alanine, or a sequence of alanine alone can be used; however, any sequence of amino acids and any length can be used that does not interfere with the binding of the steroid receptor to its response element and that does not prevent fluorescence of the fluorescent protein. Typically, a linker peptide will range from two to about ten amino acids but maybe shorter or longer. Of course, certain linker peptides maybe preferred over others, e.g., the presence of four basic amino acids in a string of six might suffice as a nuclear localization signal so as to mislocate the uninduced state of the factor. A linker peptide can be used to separate the fluorescent protein structurally from the response element and can function to allow the fluorescent protein independently of the remaining portion of the chimeric protein. An example of a

chimeric protein, which has GFP fused via five glycine-alanine repeating units to the N-terminus of the glucocorticoid receptor, is provided herein as SEQ ID NO: 2. Typically, a fluorescent protein can be fused to either the C-terminus or the N-terminus of the transcription factor; however, the preferable construction for a specific transcription factor can readily be determined. Linker peptides can readily be introduced between the two proteins in the chimeric protein by producing an nucleic acid that encodes the chimeric protein having the linker sequence between the fluorescent protein and transcription factor components.

Modifications to the fluorescent protein portion of the chimeric protein and/or the transcription factor, e.g., steroid receptor can be made. For example, a green fluorescent protein can be modified as described above. The transcription factor, for example, can be modified to increase or decrease its affinity for its binding site or to determine if a selected modification affects its binding affinity. In the case of steroid- or ligand-dependent transcription factor, the region involved in steroid- or ligand-binding can be altered to either increase or decrease the affinity to the steroid or ligand or alter the specificity of the ligand. Furthermore, other functions of the factor, such as transactivation potential, maybe modified. An example of such a modification is found in the chimeric protein having the amino acid sequence set forth in SEQ ID NO: 2, wherein the transcription factor is glucocorticoid receptor having a substitution of serine for cysteine at amino acid 656 that has a higher binding affinity for its ligand than the protein having the naturally occurring amino acid sequence. This substitution also increases the transactivation potential of the receptor, resulting in "superactivation." This cysteine 656 mutation can be utilized, for example in rat, human and mouse glucocorticoid receptor. For example, steroid receptors, or any transcription factor, can be modified in their steroid binding domains to increase affinity for steroid, thus allowing one to increase use of exogenous receptor over endogenous receptor in a cell.

The transcription factor of the chimeric protein can be derived from any selected mammal. Additionally, chimeric proteins utilizing a transcription factor from one mammal can often be used in a cell from another mammal. For example, the

glucocorticoid receptor amino acid sequence is highly conserved, particularly in the binding region among rat, human and mouse, and, for example, the rat glucocorticoid receptor binds with high affinity to the human glucocorticoid response element.

Also provided in the present invention is nucleic acid encoding a chimeric protein wherein a fluorescent protein is fused to a transcription factor. The nucleic acid encoding the chimeric protein can be any nucleic acid that functionally encodes the chimeric protein. For example, to functionally encode, *i.e.*, allow the nucleic acid to be expressed, the nucleic acid can include, for example, expression control sequences, such as an origin of replication, a promoter, an enhancer, and necessary information processing sites, such as ribosome binding sites, RNA splice sites, polyadenylation sites, and transcriptional terminator sequences. Preferred expression control sequences are promoters derived from metallothionine genes, actin genes, immunoglobulin genes, CMV, SV40, adenovirus, bovine papilloma virus, etc. A nucleic acid encoding a selected chimeric protein can readily be determined based upon the genetic code for the amino acid sequence of the selected chimeric protein, and, clearly, many nucleic acids will encode any selected chimeric protein. Modifications to the nucleic acids of the invention are also contemplated, since mutations in the steroid receptor binding can thereby be studied for binding affinity. Additionally, modifications that can be useful are modifications to the sequences controlling expression of the chimeric protein to make production of the chimeric protein inducible or repressible upon addition to the cells of the appropriate inducer or repressor. Such means are standard in the art (*see, e.g.*,. Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989). The nucleic acids can be generated by means standard in the art, such as by recombinant nucleic acid techniques, as exemplified in the examples herein, and by synthetic nucleic acid synthesis or in vitro enzymatic synthesis.

An example of a nucleic acid of the present invention is a nucleic acid encoding a chimeric protein comprising a green fluorescent protein fused via ten amino acid gly-ala linker to the N-terminus of the rat glucocorticoid receptor. One nucleic acid

encoding this nucleic acid is set forth in SEQ ID NO: 1. This nucleic acid encodes a modified *Aequorea victoria* green fluorescent protein and a modified rat glucocorticoid receptor. Another example is a nucleic acid encoding a chimeric protein comprising a modified *Aequorea victoria* green fluorescent protein fused (via ten amino acid gly-ala linker) to the N-terminus of the human estrogen receptor.

Additionally contemplated by the invention are closely related receptors and nucleic acids encoding them. Thus, provided by the invention are nucleic acids that specifically hybridize to the nucleic acids encoding the chimeric proteins under sufficient stringency conditions to selectively hybridize to the target nucleic acid. Thus, nucleic acids for use, for example, as primers and probes to detect or amplify the target nucleic acids are contemplated herein. Typically, the stringency of hybridization to achieve selective hybridization is about 5°C to 20°C below the Tm (the melting temperature at which half of the molecules dissociate from its partner). Hybridization temperatures are typically higher for DNA-RNA and RNA-RNA hybridizations. The washing temperatures can similarly be used to achieve selective stringency, as is known in the art. (Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989; Kunkel et al. *Methods Enzymol.* 1987:154:367, 1987).

The present invention provides cells containing a nucleic acid of the invention. A cell containing a nucleic acid encoding a chimeric protein typically can replicate the DNA and, further, typically can express the encoded protein. The cell can be a prokaryotic cell, particularly for the purpose of producing quantities of the nucleic acid, or a eukaryotic cell, particularly a mammalian cell. The cell is preferably a mammalian cell for the purpose of expressing the encoded protein so that the resultant produced protein has mammalian protein processing modifications. Additionally, as described above, the cell can have an array of a nucleic acid to which the encoded chimeric protein binds.

Labeled compounds and nucleic acids encoding chimeric proteins can be delivered into cells by any selected means, in particular depending upon the purpose of the delivery of the compound and the target cells. Many delivery means are well-known in the art. For example, electroporation, calcium phosphate precipitation, microinjection, cationic or anionic liposomes, and liposomes in combination with a nuclear localization signal peptide for delivery to the nucleus can be utilized, as is known in the art. In particular for transfer of a nucleic acid into a cell, to enhance transfer a cotransfection of the nucleic acid with a second nucleic acid encoding a selectable marker can be performed, and transfected cells selected for by the selectable marker. For example, the interleukin 2 receptor (IL2R) gene can be cotransfected, and selection performed by using beads having the antibody to IL2R bound to the beads to separate out transfected cells. Such methods are standard in the art.

Nucleic acids of the present invention can be used to generate transgenic animals in which the nucleic acid encoding a selected chimeric protein, such as GFP-GR for GR studies or GFP-ER for ER studies, is added to the germ line of the animals. Thus a cell of the invention containing an nucleic acid of this invention is contemplated to include a cell in a transgenic animal. With such transgenic animals, cytoplasm-to-nucleus translocation and gene targeting can be observed in any tissue of interest. Thus studies over the life cycle of the animal can be conducted, so that, for example, development and effects of environment, aging, cancer, etc. can be readily observed. Transgenic animals are generated by standard means known to those skilled in the art.

The provision of the present method to visualize physiologically relevant target sites within the eukaryotic nucleus allows one to directly observe nuclear target sites for any desired steroid- or ligand-dependent transcription factors as well as any nuclearly targeted trans-regulatory factors. Conceptually, the requirements are: 1) to tag the protein of interest with a fluorescent protein, such as the green fluorescent protein, using standard recombinant DNA techniques wherein the two cDNAs are fused in frame to each other such that introduction back into the mammalian cell would give rise to the synthesis of the chimeric protein of interest; and, 2) to create an identifiable nuclear

target site by linking together multiple copies of the potential target site to generate a large enough array as to be readily discernible when the fluorescently labeled chimeric protein interacts with the target site. Such an interaction is visible as an intense, concentrated fluorescent signal unique to the cell harboring the array and absent from the parental cell lacking the array. Agents which modify the interaction of the labeled factor with its cognate site can thus be readily screened.

The present invention further provides a method of screening for a compound that binds to a selected nucleic acid comprising:

a. contacting compound fluorescently labeled by a fluorescent protein with a cell having a plurality of copies of the nucleic acid in an array such that the nucleic acid can be directly detected when bound by fluorescently labeled compound; and

b. directly detecting the location of fluorescence within the cell, fluorescence aggregated at the site of the nucleic acid array indicating a compound that binds to the selected nucleic acid. Thus, the present method can be utilized to directly determine whether a compound binds a nucleic acid, directly or indirectly. For example, transcription factors that have been indirectly shown to affect binding can now be assessed to determine if they do bind the target DNA. If, by this method a factor labeled by a fluorescent protein, as herein described (for example, *Aequorea* GFP), is added to cells, and upon visualization of the fluorescence, the fluorescence is seen to be aggregated at the site of the target array of nucleic acid, then the factor, or a factor(s) to which the labeled factor binds, binds the nucleic acid. The nucleic acids are conceptualized as merely serving as an easily identifiable "stage" upon which the "actors" (i.e., transcription factors and associated factors) appear.

In the present inventive methods, by fluorescence aggregated at the site of the nucleic acid array is meant fluorescence aggregated at a single predominant site within the nucleus. Such aggregation can readily be detected upon excitation of the fluorescent protein. Detection, as described above, can be performed with the use of a fluorescent microscope. Magnification can be altered as desired for more or less detail in visualizing

the aggregation, such as from about 60x to about 200x, with a typical magnification being about 100x. The fluorescent protein in any herein described method can be, for example, a green fluorescent protein, as described herein.

Compounds for use in the present methods can be labeled by standard means in the art for linking a peptide to the compound. For example, when the compound to be labeled is a peptide, polypeptide or protein, a chimeric protein can be made by synthesizing a nucleic acid that encodes the chimeric protein having the fluorescent protein fused to the compound protein. Proteins can also be labeled with a fluorescent protein by a chemical bridge. Additionally, a fluorescent protein label can be placed on the compounds by non-covalent interaction, such as that displayed by steroids with their binding domain [e.g., fluorescein or rhodamine conjugated dexamethasone (available through Molecular Probes, Inc. catalog number D-1382 or D-1383) for the glucocorticoid receptor].

More than one chimeric protein, preferably each comprising a fluorescent protein that emits a different wavelength of light (for example a modification of *Aequorea* GFP that emits blue and a modification of *Aequorea* GFP that emits green), can be used simultaneously in the present inventive methods. For example, plasmids pCI-nGL2-C656G; pCI-nGL3-C656G; pCI-nGL4-C656G; pCI-nGL5-C656G; pCI-nGL7-C656G; pCI-nGL9-C656G; pCI-nGL10-C656G; pCI-nGL11-C656G, all GFP-GR having blue variants of GFP (BFP) in a humanized codon usage (improved blue variants) can be used with any GFP-GR having green GFP (such as pCI-nGFP-C656G, pCI-nGL1-C656G, pTET-nGFP-C656G and pOP-nGFP-C656G, exemplified herein) in this method.

Alternatively, two different fluorescent moieties that show distinct excitation maxima with identical emission wavelengths could also be simultaneously utilized to detect the presence of both fluorescently labeled compounds on the same DNA array. With the appropriate combination of fluorescent molecules only one excitation wavelength may be required to detect the presence of both fluorescent moieties in close

proximity through the process of fluorescent energy transfer, wherein the excitation wavelength excites one moiety which emits at the absorption wavelength of the second moiety. Such simultaneous use will allow the detection of interaction of various transcription factors and cofactors with each other and with the DNA to activate and/or repress transcription from a specific regulatory sequence.

Cells for use in the present methods are cells having a plurality of copies of the nucleic acid in an array such that the nucleic acid can be directly detected when bound by fluorescently labeled compound. Such cells are described herein and can be prepared as described herein. An example of such a cell is cell line 3134, having about two hundred copies of the MMTV LTR-Ha-v-ras- gene, each of which has four copies of the binding site for glucocorticoid receptor, in direct tandem repeats integrated into the genome of the cell. Cells further can have a nucleic acid encoding a chimeric protein comprising a fluorescent protein fused to the binding compound/transcription factor of interest, *i.e.*, the binding compound/transcription factor which may bind the nucleic acid in the array in the cell. For example, 3134 cells containing pCI-nGFP-C656G, pCI-nGFP-C656G, pTET-nGFP-C656G or pOP-nGFP-C656G are exemplified herein. Nucleic acids encoding a chimeric protein can either be integrated or not, as best suits the specific method being performed.

The present invention also provides a method of characterizing cells in which a compound fluorescently labeled by a fluorescent protein is expressed in the cell and the localization of the fluorescent protein observed for perturbation in localization of the fluorescently labeled protein in the absence or presence of signals that affect protein function (example of GFP-ER in the MCF7 and MDA-MB-231 cell line).

The present invention also provides a method of screening for a ligand that activates gene targeting of a steroid receptor in the nucleus of a mammalian cell comprising contacting the ligand with a cell having a plurality of steroid receptor response elements in an array such that the response element can be directly detected when bound by fluorescently labeled steroid receptor and the cell further comprising a

nucleic acid encoding a chimeric protein wherein a fluorescent protein is fused to the steroid receptor; and directly detecting the location of fluorescence within the cell, fluorescence aggregated at the site of the steroid receptor response element array in the nucleus indicating a ligand that activates the gene targeting of a steroid receptor in the nucleus of a mammalian cell. A ligand for any steroid receptor can be determined by this method by creating an array of the steroid receptor response element in the cell used. For example, cell line 3134 can be used to detect ligands that activate gene targeting of glucocorticoid response element, progesterone receptor, or aldosterone receptor. An example of a chimeric protein for use in this method in, for example cell line 3134, is one that has a green fluorescent protein, such as *Aequorea* fluorescent protein, fused to the N-terminus of glucocorticoid receptor, such as the chimeric protein comprising the amino acid sequence set forth in SEQ ID NO: 2.

The present invention further provides a method of screening for a ligand that activates the translocation of a steroid receptor to the nucleus or redistribution of a steroid receptor in a mammalian cell comprising contacting the ligand with a cell having a plurality of steroid receptor response elements in an array such that the response element can be directly detected when bound by fluorescently labeled steroid receptor and the cell further comprising a nucleic acid encoding a chimeric protein wherein a fluorescent protein is fused to the steroid receptor; and directly detecting the location of fluorescence within the cell, change in the location of the fluorescence *(e.g.*, cytoplasm to nucleus in the case of the glucocorticoid receptor or redistribution within each cellular compartment; *e.g.*, for primarily nuclear receptors, such as the progesterone receptor, aggregation on the MMTV LTR-array in the 3134 cell) could indicate a potential ligand of the receptor. In the present method, redistribution of the receptor can be directly observed. Also, in the case of the glucocorticoid receptor, translocation to the nucleus, even in the absence of binding to the nuclear DNA, can be seen, and importantly, one can see if a ligand causes only translocation to the nucleus (by location of fluorescence primarily in the nucleus, but in a diffuse or reticular, rather than aggregated, pattern) or causes both translocation to the nucleus and binding to nuclear DNA (by location in the nucleus aggregated primarily at a site). In the case of estrogen

receptor which has been partially activated due to trace estrogenic substances present in the culturing media, the receptor is nuclearly localized but shows dramatically different intranuclear distribution in two human breast cancer cell lines. In the case of MCF7 cell, a human adenocarcinona breast cell line which contains endogenous estrogen receptor and shows hormone dependent growth, the GFP-tagged estrogen receptor is collected upon nuclear structures and shows focal accumulation patterns. In contrast, the MDA-MB-231 cell, a human adenocarcinoma breast cell line which lacks endogenous estrogen receptor and shows hormone-independent growth, the GFP-tagged estrogen receptor under the same culturing condition is extremely diffuse. These two dramatic differences in the nuclear localization patterns in two different human breast cancer cell lines suggests potential usefulness of GFP-ER as a diagnostic reagent for characterizing different human breast cancer cells as well as characterizing the progression of human breast cancer. These differences in GFP-ER localization patterns also suggest an additional requirement for celllular components in permitting the targeting of the estrogen receptor onto nuclear structures; these cellular components maybe absent as the cell progresses from hormone-dependent to a hormone-independent stage in the progression of human breast cancer.

The present invention additionally provides a method of detecting in a biological sample the presence of an agonist of a steroid receptor comprising contacting the sample with a cell having a plurality of steroid receptor response elements in an array such that the response element can be directly detected when bound by fluorescently labeled steroid receptor and the cell further comprising a nucleic acid encoding a chimeric protein wherein a fluorescent protein is fused to the steroid receptor; and directly detecting the location of fluorescence within the cell, the location of fluorescence aggregated at the site of the steroid receptor response element array in the nucleus indicating the presence of an agonist of the steroid receptor in the sample.

Biological samples can include any relevant sample from the body, such as blood, plasma, urine and saliva.

The present invention also provides a method of detecting in a biological sample the presence of an antagonist of a steroid receptor comprising contacting the sample and both agonist and antagonist of the steroid receptor with a cell having a plurality of steroid receptor response elements in an array such that the response element can be directly detected when bound by fluorescently labeled steroid receptor and the cell further comprising a nucleic acid encoding a chimeric protein wherein a fluorescent protein is fused to the steroid receptor; and directly detecting the location of fluorescence within the cell, the absence of fluorescence substantially aggregated at the site of the steroid receptor response element array in the nucleus indicating the presence of an antagonist of the steroid receptor in the sample.

Also provided is a method of monitoring the level of an agonist of a steroid receptor in a subject comprising periodically obtaining a biological sample from the subject, contacting the sample with a cell having a plurality of steroid receptor response elements in an array such that the response element can be directly detected when bound by fluorescently labeled steroid receptor and the cell further comprising a nucleic acid encoding a chimeric protein wherein a fluorescent protein is fused to the steroid receptor; and directly detecting the location of fluorescence within the cell, a decrease in fluorescence aggregated at the site of the steroid receptor response element in the nucleus in a later-obtained sample relative to an earlier-obtained sample indicating a decrease in level of the steroid agonist of the steroid receptor in the sample and an increase in fluorescence aggregated at the site of the steroid receptor response element in the nucleus in a later-obtained sample relative to an earlier-obtained sample indicating an increase in level of the steroid agonist of the steroid receptor in the sample.

Further, provided by the present invention is a method of monitoring the balance between levels of an agonist of a steroid receptor and an antagonist of the steroid receptor in a subject comprising periodically obtaining a biological sample from the

subject, contacting the sample with a cell having a plurality of steroid receptor response elements in an array such that the response element can be directly detected when bound by fluorescently labeled steroid receptor and the cell further comprising a nucleic acid encoding a chimeric protein wherein a fluorescent protein is fused to the steroid receptor; and directly detecting the location of fluorescence within the cell, an increase in fluorescence aggregated at the site of the steroid receptor response element in the nucleus in a later-obtained sample relative to an earlier-obtained sample indicating an increase in level of the steroid agonist relative to level of the steroid antagonist in the sample, and a decrease in fluorescence aggregated at the site of the steroid receptor response element in the nucleus in a later-obtained sample relative to an earlier-obtained sample indicating an increase in level of the steroid antagonist of the steroid receptor relative to level of the steroid agonist in the sample.

The present invention also provides a method of determining an effective dosage of a steroid receptor agonist in a subject comprising transferring into a set of cells from the patient a nucleic acid encoding a chimeric protein comprising a fluorescent protein fused to a steroid receptor; contacting the cells in the set with one of a selected range of dosages of the steroid agonist; and directly detecting location of fluorescence in the set of cells, a dosage capable of locating fluorescence substantially in the nucleus indicating an effective dosage of steroid receptor agonist.

Further provided by the present invention is a method of determining an effective dosage of a steroid receptor agonist to maintain steroid receptor activation for a selected period of time in a subject comprising administering to the subject a dosage of the steroid receptor agonist; periodically obtaining a biological sample from the subject; contacting the sample with a mammalian cell having a plurality of steroid receptor response elements in an array such that the response element can be directly detected when bound by fluorescently labeled steroid receptor and the cell further comprising a nucleic acid encoding a chimeric protein wherein a fluorescent protein is fused to the steroid receptor; and directly detecting the location of fluorescence within the cell, a dosage that maintains the location of fluorescence at the site of the steroid receptor

response element array in the nucleus for the selected period of time indicating an effective dosage.

The present invention additionally provides a method of detecting a defect in a response pathway of a steroid receptor in a subject comprising transferring into a cell from the subject a nucleic acid functionally encoding a chimeric protein comprising a fluorescent protein fused to the steroid receptor and detecting the location of fluorescence within the cell as compared to the location of fluorescence within a normal, control cell transfected with the nucleic acid, a difference in location of fluorescence within the cell of the subject as compared to location of fluorescence within the normal, control cell indicating a defect in the response pathway of the steroid receptor.

The present invention also provides a method of determining whether a defect in a response pathway of a steroid receptor in a subject is in translocation of the steroid receptor to a cell nucleus, comprising transferring into a cell from the subject having the defect a nucleic acid functionally encoding a chimeric protein comprising a fluorescent protein fused to the steroid receptor and detecting the location of fluorescence within the cell, the location of fluorescence substantially in the cytoplasm of the cell indicating the defect is in translocation of the steroid receptor to the nucleus.

The present invention also provides a method of characterizing a ligand's effect on cellular localization of a compound to which the ligand binds in a cell comprising contacting the ligand with a cell having the compound fluorescently labeled by a fluorescent protein and directly detecting the location of fluorescence within the cell, the location of fluorescence in the cell indicating the localization effect of the ligand on the compound. Compounds can be, *e.g.*, steroid receptors, transcription factors and the like. For example, the examples provide characterization of localization of GR in response to two ligands, dexamethasone and RU486, and characterization of the localization of ER in response to ligands, agonist beta-estradiol or anti-estrogens, 4-hydroxytamoxifen or ICI164384." By this method, the ligands triggering binding of so-called "orphan receptors" to their binding site(s) can be discovered.

Additionally provided is a method of determining a binding site for a DNA-binding protein comprising contacting the DNA-binding protein fluorescently labeled by a fluorescent protein with a cell having a plurality of copies of a nucleic acid having a putative binding site in an array such that the putative binding site can be directly visualized when bound by the fluorescently labeled DNA-binding protein, and directly detecting the location of fluorescence within the cell, the presence of fluorescence aggregated at the putative binding site indicating a binding site to which the DNA-binding protein binds. The absence of fluorescence aggregated at the putative binding site can suggest a binding site to which the DNA-binding protein does not significantly bind.

the location of fluorescence in the cell indicating the localization effect of the ligand on the compound.

Also provided by the present invention is a method for screening for gene-specific combinations of compounds that bind the gene specifically, comprising contacting (a) a first compound labeled by a fluorescent protein emitting a first spectrum of light and a second compound labeled by a fluorescent protein emitting a second spectrum of light with (b) a cell having a plurality of copies of the regulatory region of the gene in an array such that the regulatory region can be directly detected when bound by compound labeled by a fluorescent protein. Fluorescence for the first and second spectrum is then localized. Aggregation of fluorescence of both the first and second spectrum at the site of the regulatory region array would indicate a combination of compounds that binds the gene specifically; the location of only one spectrum aggregated at the array would indicate that only the corresponding compound binds the gene of interest directly. By using a combination of screens, compounds that bind the specific DNA both directly and indirectly can be determined for a gene of interest. Such gene-specific combinations of compounds can be used to develop gene-specific drugs that interfere with transcription activators in a selective manner. This method is based on the fact that each transcription factor, cofactor, etc. affects many genes, but for each gene there is likely only one combination of factors and cofactors that activates/represses it. Therefore, once, by this screening method, it has been determined

which combination of factors causes activation or repression of a specific gene, then a combination of drugs, to affect all relevant factors for that specific gene, can be administered to selectively activate/repress that gene. Thus a combination of drugs can ultimately be used to activate or repress the selected gene.

Statement Concerning Utility

The present invention provides methods for directly detecting the binding of compounds to nucleic acids. The present invention allows a simple and straightforward manner in which direct interaction between a sequence-specific DNA binding protein or its co-factor and its putative regulatory site in the *in vivo* genomic context can be addressed. For example, GFP-steroid fusion proteins and cell lines containing receptor binding sites (response elements) in multimerized arrays are provided for direct visualization of *in vivo* gene targeting. These reagents provide a simple, rapid, straightforward, sensitive, and biologically relevant assay for each target nucleic acid and binding compound. These reagents can in turn be used for several medically important applications, including diagnostic tests for concentration of cognate ligand in clinical samples (urine, saliva, blood, etc.). Additionally, the reagents can be used for direct tests for defects in steroid pathways in subjects. The reagents further can be used for screening of chemical banks for compounds with ligand agonist activity for each receptor, and development of drugs based on these activities. Furthermore, using the fluorescent protein-steroid receptor fusions, colocalization of a selected receptor with any other cofactor that may be recruited to the chromosome template can be evaluated. Using a separate tag (different color) for the DNA target, the steroid receptor can be fused to any factor that may be recruited by the steroid receptor and determine if the factor colocalizes on the DNA target when the steroid receptor is activated, allowing for a direct test for defects in factor colocalization in human disease/syndromes. This ability to observe direct interaction of any trans-regulatory factor or co-factor and its regulatory site *in vivo* provides a screening method for useful and novel drugs, directed against trans-regulators, and the development of gene-specific multi-drug therapies. Using the glucocorticoid receptor as an example and known ligands of the receptor (*e.g.*, agonist dexamethasone and antagonist RU486), the validity of the methodology is

herein demonstrated. In a specific example, the ability to monitor effective hormone concentration by this novel methodology in real time can lead to the development of diagnostic kits that can be used to properly gauge the required amounts of hormone administered to patients requiring long-term or short-term hormone treatment. In addition, in the case of those trans-acting factors regulated at the level of the nuclear/cytoplasmic and cytoplasmic/nuclear translocation step, diseases arising from such a failure can be directly diagnosed by fusion of the regulatory molecule to a fluorescent moiety using standard recombinant DNA techniques. With this simple, inventive methodology, novel classes of drugs directed not only against members of the steroid-ligand-dependent transcription factors but to new classes of drugs that target other transcription factors or their co-factors can be screened. Furthermore, by using combination of drugs which target certain trans-regulatory factors either specifically or selectively, a gene-specific based drug therapy regimen can be created. This multi-drug therapy designed against a certain critical gene implicated for a particular human disease would be tailored to affect the activity of the trans-regulatory factors all of which act synergistically to regulate the transcription of the gene implicated in the disease. Furthermore, the reagents allow the development of transgenic animals containing fusion proteins such as each of the GFP-receptor fusions that can provide a unique tool to study subcellular distribution of the receptors in all tissues of the animal, and the effect of pharmacologic agents on function of each of the receptors in the various tissues. Numerous other utilities will be apparent to the skilled artisan in light of the present invention.

EXAMPLES

The present invention is more particularly described in the following examples which are intended as illustrative only since numerous modifications and variations therein will be apparent to those skilled in the art.

Plasmids.

pCI-nGFP-C656G: Plasmid pCI-nGFP-C656G was derived from pCI-nH6HA-C656G (Smith et al., submitted) and pZA69 (a kind gift from Mike Moser and Ravi Dahr). pZA69 contains a BspHI fragment of pZA66, a plasmid containing S65T GFP (Tsien, R.Y. (1995) *Nature* **373**,663-664) with the internal NcoI site removed by a silent mutation. The pCI-nH6HA-C656G DNA expresses the rat glucocorticoid receptor with the C656G mutation (kindly provided by S. Simons, Jr.) under the control of the CMV promoter/enhancer, and is tagged at the N-terminus with $(his)_6$ and hemagglutinin epitope recognized by monoclonal antibody 12CA5 (Niman, H. L., Houghten, R. A., Walker, L. E., Reisfeld, R. A., Wilson, I. A., Hogle, J. M. & Lerner, R. A. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80**, 4949-4953); this DNA was cleaved at a unique site with PvuII, separating the two tags from the rest of the glucocorticoid receptor. A 768 bp DNA fragment containing the GFP cDNA with the S65T mutation was inserted at the PvuII site. This GFP DNA fragment is obtained by digesting pZA69 with BglI, attaching a BglI hairpin linker with the sequence (5'-GCGCGCTGATCAGAATTCCTTTTAGGAATTCTGATCAGCGCGCTGA-3') (SEQ ID NO:3), recutting the resulting DNA with BclI and XhoI, and then filling-in with the large fragment of DNA polymerase (Klenow) to create a 768 bp blunt-end fragment.

pCI-nGL1-C656G: The GFP (S65T) variant with the jellyfish codon usage in the mammalian expression GFP-GR vector, pCI-nGFP-C656G, has been replaced with a humanized codon usage from the plasmid, pGreenLantern-1 (LifeTechnologies, Inc., Gaithersburg, MD, catalog number 10642-015), for improved translation in mammalian cells [this humanized GFP is also approved for in vitro diagnostic use by LifeTechnologies] to generate an improved mammalian expression GFP-GR plasmid, pCI-nGL1-C656G.

pCI-nGL2-C656G; pCI-nGL3-C656G; pCI-nGL4-C656G; pCI-nGL5-C656G; pCI-nGL7-C656G; pCI-nGL9-C656G; pCI-nGL10-C656G; pCI-nGL11-C656G: Blue variants of GFP (BFP) in a humanized codon usage (improved blue variants) have been generated which are fused to the rat glucocorticoid receptor (C656G). Each has a

different chromophore. Site-directed mutagenesis of the GFP element of pCI-nGL1-C656G (which has humanized GFP) was performed to introduce known chromophore-altering mutations into the GFP element. Site-directed mutagenesis was performed using Chameleon™ double-stranded site-directed mutagenesis kit (Stratagene, catalog number 200509). These new fusion plasmids were then expressed in 1471.1 cells. Fluorescent spectrophotometry indicates the chromophores are present and that the GFP-GR fusion is intact. Additional mutants, such as those producing a longer wavelength chromophore, *e.g.* red chromophores, can be made by the same methods.

GFP-ER plasmid: The glucocorticoid receptor portion of pCI-nGL1-C656G DNA has been replaced with a human estrogen receptor (ER). ER binds to an ER response element in cells. The GFP-ER was found to be functional in both transcriptional activation as well as proper subcellular localization in several cell lines, as described below. GFP-ER plasmid is made using the same site-directed mutagenesis as for preparing blue variant plasmids. Briefly, a cDNA encoding ER is inserted into pCI-nGL1-C656G in place of the GR cDNA. An ER cDNA (Green, Stephen, *et al.*, *Nature* 320:134-139 (1986) (GenBank accession number X03635) (note: this sequence has a Gly_{400} to Val_{400} mutation); Greene, Geoffrey L., *et al. Science* 231:1150-1154 (1986) (GenBank accession number M12674 (having Gly_{400} to Val_{400} mutation)); Pfeffer, U. *Cancer Res.* 53:741-743 (1993) (GenBank accession number X73067) (ER fragment having correct Gly_{400} region coding sequence)) is mutagenized to create a MluI site at the start point of translation of ER and a SalI site in the 3' untranslated region of the cDNA (alternatively, the ER sequence can be generated by PCR). pCI-nGL1-C656G has a unique BssHII site after the (gly-ala)5 linker and a SalI site after the GR portion of the plasmid DNA. The DNA cut end made by MluI is complementary to the end made by the BssHII. Therefore, the GR cassette can be removed by BssHII/SalI digestion and the ER cassette (released by MluI/SalI digestion) subcloned into the remaining vector at the BssHI and SalI cut ends.

pOP-nGFP-C656G: The original cDNA encoding GFP-GR from the plasmid, pCI-nGFP-C656G, has been subcloned into a tetracycline-regulatable mammalian

expression vector, pTET-Splice (LifeTechnologies, catalog number 10583-011), to give pTET-nGFP-C656G, and a lac-regulatable mammalian expression vector, pOPRSVI CAT (Stratagene Cloning Systems, catalog number 217450) to give pOP-nGFP-C656G.

Other plasmids used in this study are: pLTRLuc (full-length MMTV LTR driving the expression of a luciferase gene) (Lefebvre, P., Berard, D. S., Cordingley, M. G. & Hager, G. L. (1991) *Mol. Cell. Biol.* **11(5)**, 2529-2537), pCMVIL2R (IL2R expression plasmid) (Giordano, T., Howard, T. H., Coleman, J., Sakamoto, K. & Howard, B. H. (1991) *Exp Cell Res.* **192**, 193-197), and pUC18 (Life Technologies, Inc.).

GFP-fusion plasmids of the invention are tested for expression and subcellular localization by transfection into several mammalian cells. For example, GFP-GR plasmids were analyzed in C127 (mouse) cells, HeLa (human cervical cancer) cells, and MCF7 (human breast adenocarcinoma) cells (ATCC accession number HTB22). GFP-ER plasmids were analyzed in 1471.1 cells, C127 (mouse) cells, MCF7 (human breast adenocarcinoma) cells, and MDA-MB-231 (human breast adenocarcinoma) cells (ATCC accession number HTB 26). Localization is observed in the absence of added hormone in either 5% or 10% charcoal-stripped fetal calf serum. Cells having GFP-fusion plasmids are then treated with a selected ligand, and subcellular localization and quantitative observations are made.

Cell Line 1471.1 and derivatives

Cell line 1471.1 contains multiple copies of a BPV MMTV-LTR-chloramphenicol acetyltransferase (CAT) reporter gene fusion introduced in the murine adenocarcinoma C127 cell (Archer, T. K., Cordingley, M. G., Marsaud, V., Richard-Foy, H. & Hager, G. L. (1989) in *Proceedings: Second International CBT Symposium on the Steroid/Thyroid Receptor Family and Gene Regulation*, eds. Gustafsson, J. A., Eriksson, H. & Carlstedt-Duke, J. (Birkhauser Verlag AG, Berlin), pp. 221-238).

Derivatives of 1471.1 cells which contain over a thousand copies of the MMTV LTR-CAT have also been generated with the tetracycline- and lac-regulatable GFP-GR (pTET-nGFP-C656G and pOP-nGFP-C656G), *e.g.*, cell line 3677. In these derivative cell lines, GFP-GR expression occurs upon tetracycline withdrawal from 5 ug/ml or upon induction with IPTG using standard procedures and as recommended by the manufacturer of these inducible systems.

Transfection. Plasmid DNA was transiently introduced into 1471.1 cells either by calcium phosphate coprecipitation using a BES-based buffer (Chen, C. & Okayama, H. (1987) *Mol. Cell Biol.* 7, 2745-2752) or by electroporation. For calcium phosphate coprecipitation, semi-confluent cells maintained in Dulbecco's modified eagle media (DMEM; Life Technologies, Inc.) supplemented with 10% fetal calf serum (FCS; Life Technologies, Inc.), 2 mM glutamine, and 50 mg/ml gentamicin sulfate were trypsinized, washed, resuspended at $7x10^4$ cells/ml in DMEM supplemented as above except FCS was treated with charcoal/dextran-treated fetal bovine serum (Hyclone Laboratories, Inc.), dispensed as 1 ml into 2x2 cm^2 Lab-Tek Chamber Slide (Nunc, Inc.) or as 10 ml into a 100 mm petri dish layered with 24.5 mm-diameter Dvorak-Stotler coverslips (Nicholson Precision Instrument). Cells were grown overnight in a 37°C humidified incubator with 5% CO_2. The following morning, media was replaced with fresh supplemented DMEM containing dextran/charcoal treated FCS, and in the afternoon, cells were transfected with 1 ml of transfection mixture containing 20 µg plasmid DNA (as indicated in the appropriate figure legend) per 10 ml of cells, essentially as described (Chen, C. & Okayama, H. (1987) *Mol. Cell Biol.* 7, 2745-2752). Cells were left overnight in a 37°C humidified incubator with 2.9% CO_2. About 12-16 hours after transfection, media was replaced, and the cells were allowed to recover for two hours before further treatment and imaging. For calcium depletion experiments, cells were electroporated with 5-20 µg pCI-nGFP-C656G DNA for $2X10^7$ cells in 0.2 ml cold DMEM at 250 V and 800 µF, left to recover on ice for 5 minutes, and then diluted in DMEM supplemented with dextran/charcoal treated FCS before plating. Cells were then grown for 12 to 16 hours in a 37°C humidified incubator at 5% CO_2. Before treatment and imaging, cells were fed with fresh media.

Enrichment of Transfected Cells and Analysis of Cytosolic Extracts. Cells that took up exogenous DNA were enriched by cotransfection with pCMVIL2R, an IL2R (interleukin 2 receptor) expression plasmid, and selection for $IL2R^+$ cells using magnetic beads (Dynal) coated with mouse anti-human IL2R antibody (Boehringer Manneheim, clone 3G10), as described (Giordano, T., Howard, T. H., Coleman, J., Sakamoto, K. & Howard, B. H. (1991) *Exp Cell Res.* **192**, 193-197). Extracts from the $IL2R^+$ and $IL2R^-$ cells were made by three cycles of freezing and thawing of the cell suspension in either 100 mM sodium phosphate (pH 7.8) with 1 mM DTT or 250 mM Tris-HCl (pH 7.8). After clarifying the lysate, extracts made with the phosphate buffer was used to assay for the amount of luciferase activity in a MicroLumat LB96P as recommended by the manufacture, EG&G Berthold. For the Tris-buffered extract, CAT activity was assayed as described (Gorman, C. M., Moffat, L. F. & Howard, B. H. (1982) *Mol. Cell Biol.* **2**, 1044-1051). Protein concentration was determined by the method of Bradford using the Bio-rad Protein Assay reagent (Bio-rad Laboratories, Inc.).

Determination of Intracellular Calcium. Intracellular free calcium concentrations were determined in single cells by measuring the signal from the calcium sensitive indicator Fura-2, according to Tsien and Harootunian (Tsien, R. Y. & Harootunian, A. T. (1990) *Cell Calcium* **11**, 93-109). Briefly, cells were cultured on cover slips and electroporated with the GFP-GR chimera one day before microscopy. In preparation for imaging, cells were treated for 30 min with either assay buffer (Hank's balanced salt solution without phenol red, with 2 mg/ml glucose and 1 mg/ml BSA, containing 3 mM Ca^{++}) or with calcium-free buffer (Eagle's No. 2 medium without calcium, containing 1 mg/ml BSA, 5 mM EGTA, 5 µM thapsigargin, 2 µM ionomycin). The cells were then loaded with 5 µmol/L Fura-2-AM (from Molecular Probes Inc.) and 0.02% pluronic F-127 with either calcium-free or calcium supplemented media (30 min at room temperature, washed three times, then incubated for 15 min at 37°C). After loading, cells were placed into a Dvorak-Stotler chamber (inner volume 224 µl) and perfused at 37°C with either calcium-containing or calcium-free media. Intracellular

calcium content was measured in three independent experiments--at least 20 cells in each experiment. Ratio imaging was performed using Image 1 software (Universal Imaging Corp.) running on an IBM PC, using 340 nm and 380 nm excitation, 510 nm emission, and 490 nm dichroic barrier filters, a Zeiss Photomicroscope III microscope, enclosed into a temperature controlled incubator, and an intensified (Videoscope) CCD camera (Dage 72), and optical disc recorder (Panasonic). The system was calibrated for $[Ca^{2+}]_i$ measurement using Fura-2 pentapotassium salt and calibration buffer kit from Molecular Probes Inc. Intracellular free calcium concentrations in cells with calcium supplemented buffer were 350±183 nM, while in calcium-free buffer 60±11 nM.

Image Acquisition and Analysis. For time course studies, cells were placed into a Dvorak-Stotler chamber (inner volume 224 µl) and perfused at 37°C with assay buffer for 3 min, then with the same buffer containing 1 nM dexamethasone (dex) for 2 hours at 10 ml/hour flow rates. Samples were evaluated using a Zeiss Axiovert 10 microscope surrounded by an incubator and equipped for epifluorescence with illumination from XBO burner, 480 nm excitation and 535 nm emission and 505 nm dichromatic barrier filters (from Chroma Technology Corp.). Images were acquired every 15 seconds with a high resolution, cooled CCD camera equipped with an electromechanical shutter (Photometrics p200). Images were collected on Silicon Graphics workstations (4D310-VGX), using custom software, incorporating functions from a vendor supplied library (G.W. Hannaway & Assoc.). Experiments requiring real time image acquisition were performed on the imaging system described for the intracellular calcium measurements.

Confocal laser scanning microscopy was carried out on a Nikon Optiphot microscope equipped with BioRad MRC-600 confocal laser scanning unit, with fluorescent excitation produced by the 488 nm line of a krypton-argon laser, and using a Fluor 100/1.3 oil phase objective. From living cells expressing GFP-GR, serial 0.5 µm optical sections were collected and the digitized images were imported into a Silicon Graphics Indigo 2 workstation. Three dimensional image rendering, analysis and reconstruction was carried out with the ANALYZE software from the Mayo Clinic.

Tagging of a Highly Dexamethasone-Sensitive Form of GR with a Highly Fluorescent Variant of GFP. To develop a highly efficient, fluorescent version of the glucocorticoid receptor, we generated a GFP-GR chimera in which cDNA encoding a 27 kDa GFP variant is fused in frame to the second amino acid of a rat glucocorticoid receptor (Fig. 3A). The GFP variant contains a serine to threonine substitution at amino acid 65 (S65T mutation) from the jellyfish *Aequorea victoria,* which increases the efficiency of formation of the GFP chromophore by accelerating the rate of oxidation required for chromophore generation. In addition, the resulting chromophore is six-fold more fluorescent than the wild-type GFP (Heim, R., Cubitt, A. B. & Tsien, R. Y. (1995) *Nature* **373**, 663-664), making the use of this chromophore perhaps the most sensitive method for labelling proteins (Wang, S. & Hazelrigg, T. (1994) *Nature* **369**, 400-403). Additionally, the chromophore is formed faster, potentially explaining why expression at 37°C in mammalian cells is achieved herein, contrary to reports that the GFP chromophore does not form at a relatively high temperature of 37°C (Ogawa, *et al.: Proc. Natl. Acad. Sci. USA* 92:11899-11903 (1995)).

Since glucocorticoid receptor is ubiquitously present in all mouse cells and selective activation of the tagged receptor is required to assess the functionality of the receptor, we therefore used a glucocorticoid receptor having a higher affinity for its ligand than the endogenous receptor. To this end, S65T GFP was fused to a rat glucocorticoid receptor that contains a cysteine to glycine mutation at position 656 of the steroid binding domain (Chakraborti, P. K., Garabedian, M. J., Yamamoto, K. R. & Simons, S. S. J. (1991) *J. Biol. Chem.* **266**, 22075-22078). This point mutation, C656G, increases the affinity of the receptor ten-fold for its ligand. A dose response curve shows complete activation of GFP-GR at 1 nM dexamethasone and half maximum at 0.1 nM; the endogenous mouse receptor is fully activated at 100 nM dexamethasone with the half maximal stimulation at 10 nM. Thus, presence of the C656G mutation permits selective activation of the transfected chimeric receptor without activation of the endogenous receptor.

Transcriptional Competence of GFP-GR. When the plasmid encoding this chimera, pCI-nGFP-C656G, is introduced into cultured mouse cells, a fusion polypeptide with the predicted molecular weight of 118 kDa is produced. When these cells are stimulated with 1 nM dexamethasone, a co-transfected reporter construct (pLTRLuc) containing the luciferase reporter gene under the control of the mouse mammary tumor virus promoter (MMTV LTR) is activated (Fig. 3B). In addition, 1 nM dexamethasone-treated cells show accumulation of luciferase activity dependent on the amount of GFP-GR expression plasmid included in the transfection. In the absence of any GFP-GR expression plasmid, no significant luciferase activity accumulated in the 1 nM dexamethasone-treated cells, indicating that 1 nM dexamethasone activated the GFP-GR chimeric protein but not the endogenous GR. With increasing amount of the GFP-GR expression plasmid, luciferase activity in the 1 nM dexamethasone-treated IL2R$^+$ cells reaches the same level as that in 100 nM dexamethasone-treated IL2R$^+$ cells lacking GFP-GR expression plasmid (compare 2 μg pCI-nGFP-C656G, 1 nM dex with 0 μg pCI-nGFP-C656G, 100 nM dex). Since the 100 nM dexamethasone treatment gives complete activation of the endogenous GR in the latter case, we conclude that the GFP-GR chimeric receptor is fully functional in dexamethasone-mediated transcriptional activation of the transiently introduced reporter plasmid DNA.

Furthermore, derivative cell lines of 1471.1 which contain over a thousand copies of the MMTV LTR-CAT have also been generated with the tetracycline- and lac-regulatable GFP-GR (pTET-nGFP-C656G and pOP-nGFP-C656G), allowing GFP-GR expression upon tetracycline withdrawal from 5 ug/ml or upon induction with IPTG using standard procedures and as recommended by the manufacturer of these inducible systems. In these cells, acceptable levels of GFP-GR is reached after overnight induction of the regulatable promoters.

Assay for Ligand Effects: GFP-fusion plasmids of the invention are assayed for effects of a selected ligand on subcellular localization by transfection into selected cells. C127 (mouse) cells, HeLa (human cervical cancer) cells, and MCF7 (human breast adenocarcinoma) cells (ATCC accession number HTB22) were each transfected with

GFP-GR plasmids and with GFP-ER plasmids. Localization is first observed in the absence of added hormone in either 5% or 10% charcoal-stripped fetal calf serum. Cells having GFP-fusion plasmids are then treated with a selected ligand, and subcellular localization and quantitative observations are made.

GFP-GR Results: GR is observed to be localized to the cytoplasm in the absence of added ligand. However, when cells having the GFP-GR plasmids were treated with dexamethasone, foci of fluorescent signal (massive variation in the peak and valley of fluorescent signal) were observed in the nucleus (*see, e.g.*, FIG 4(A)). When these cells were treated with RU486, the fluorescent signal is near uniform, with a matrix-like appearance where fluorescent signal appears (*see, e.g.*, FIG 4(B)).

GFP-ER Results: ER is observed to be localized to the nucleus in the absence of added estrogen hormone. When a hormone-dependent, estrogen receptor positive human breast cancer cell line, MCF7, is transfected with GFP-ER expression plasmid, expression
of GFP-ER results in nuclear signal that is structured with peaks and valley of concentration of signals (FIG. 5(A)). Although the GFP-ER is already partially activated due to trace estrogenic substances present in the culturing medium, addition of agonist beta-estradiol leads to further concentration of nuclear signals on nuclear structures. In the case of a hormone-independent, estrogen receptor negative human breast cancer cell line, MDA-MB-231, the expressed GFP-ER in the absence of added ligand shows a diffuse, fuzzy pattern, with only hints of attachments to structures in the nucleus.

These results provide an example of use of the present method as a diagnostic for missing, dysfunctional or nonfunctional components in any selected cell. For example, upon determining that a ligand, such as estrogen for ER binding to its response element, is present in a cancer cell and used for growth or maintenance of the cell, one can treat the cancer by administering a compound to deprive the cell of that ligand, such as by administering anti-estrogen to a breast cancer showing the same results as MCF7

(hormone-dependent) to reduce growth of the cells. Similarly, such a diagnostic can tell one if such a treatment in another cell would be futile because the cell is not dependent upon that ligand.

These results also demonstrate that the present method can be used as a screen to classify cell types for the ability (or lack of ability) to target nuclear structures, to traffic compounds in a particular pattern, etc. This also provides information for selecting treatment regimens for various diseases or disorders, based on activating, inactivating or altering the function of ligands in the cell.

Transcriptional activation of the MMTV LTR target genes by GFP-GR shows a ligand specificity characteristic of glucocorticoid receptor. When activated by dexamethasone, GFP-GR is competent to induce not only the transiently introduced MMTV LTR-luciferase reporter DNA, as mentioned above (Fig. 3B), but also the multi-copy MMTV LTR-CAT reporter genes present in 1471.1 cells (1 nM and 10 nM dex, Fig. 3C). In contrast, treatment with 10 nM RU486, an antagonist with little GR agonist activity, or progesterone, a poor agonist, results in little activation of the MMTV LTR-CAT reporter; 17-β-estradiol, a steroid that shows no affinity for GR, fails to activate the LTR. Thus, the ligand specificity of GR in the transcriptional activation of the MMTV LTR is maintained in GFP-GR expressing cells.

Visualization of GFP-GR Cytoplasm-to-Nuclear Translocation in a Single Metabolically Active Cell. Because the S65T variant of the GFP chromophore is resistant to photobleaching (Heim, R., Cubitt, A. B. & Tsien, R. Y. (1995) *Nature* **373**, 663-664), it was possible to use confocal and time-lapse video microscopy to observe GFP-GR over extended periods. Using computer controlled high resolution video and confocal laser scanning microscopy, we examined transfected samples for subcellular localization of the chimeric GFP-GR protein. We observed significant fluorescence in the cytoplasm of about ten percent of total cells, approximately the fraction that typically acquires transfected DNA. Thus, the GFP was functional as a chromophore in a majority of the expressing cells in this mammalian system.

Upon exposure to dexamethasone, translocation of GFP-GR occurs in 100% of fluorescing cells, with the rate of cytoplasm-to-nuclear translocation dependent on the concentration of hormone. At 10 nM, complete translocation was induced within 10 min at 37°C, with half maximal nuclear accumulation at 5 min; this rate is consistent with previous findings (Picard, D. & Yamamoto, K. R. (1987) *EMBO. J.* **6**, 3333-3340). The rate of translocation is decreased with 1 nM dexamethasone (complete translocation over 30 min with half maximum at 9-10 min) and further reduced with 0.1 nM dexamethasone (complete translocation within 2 hours with half maximum at 1 hour).

Analysis of a time-lapse series revealed that GFP-GR accumulated along fibrillar structures and in the perinuclear region very rapidly after hormone addition, probably within seconds. Murine adenocarcinoma cells were cultured on cover slips and transfected with GFP-GR fusion chimera one day before microscopy. Cells were placed into a Dvorak-Stotler chamber and perfused at 37°C with assay buffer for 3 min, then with buffer containing 1 nM dexamethasone for 2 hrs. With real-time imaging, perinuclear accumulation was observed in a pulsatile pattern with 1-2 second intervals between brightness changes. GFP-GR accumulation was more intense along fibrillar structures in the perinuclear region. After 3 min with 1 nM dexamethasone, GFP-GR was noticeably present in the nucleus, but not in the nucleoli. When approximately 1/3 of the protein had been translocated (9-10 min), a punctate pattern appeared, and translocation was complete after 30 min. During translocation, the cells frequently became rounded, and moved along the long axis of the cell. We observed reduction of the cell surface, as well as the nuclear volume during the translocation. One hour after hormone treatment, the cells reattach and regain a more flattened shape.

Ligand Specificity of Cytoplasm-to-Nuclear Translocation. Cells were treated with buffer (A), 10 nM 17-β-estradiol (B), 10 nM dexamethasone (C), or 10 nM RU486 (D) for 30 min at 37°C. At the end of hormone treatment, images from living cells expressing GFP-GR were visualized with confocal laser scanning microscope as described above. While complete translocation of GFP-GR was observed in all fluorescing cells treated with dexamethasone, other classes of steroid hormones induced

GFP-GR translocation to varying extent reflective of the affinity for GR. The glucocorticoid antagonist RU486, known to have a high affinity for GR (Chakraborti, P. K., Garabedian, M. J., Yamamoto, K. R. & Simons, S. S. J. (1991) *J. Biol. Chem.* **266**, 22075-22078), was as potent as dexamethasone for induction of translocation. Progesterone, a weak GR agonist, required a concentration 100-fold higher than dexamethasone for translocation; however, approximately 1/2 of the GFP-GR remained in the cytoplasm. In contrast, 17-β-estradiol, a steroid hormone that does not bind GR, did not cause intranuclear GFP-GR accumulation (10 nM). Thus, GFP-GR maintained ligand-dependent cytoplasm-to-nuclear translocation, with analog specificity identical to that for the untagged GR with the C656G point mutation (Chakraborti, P. K., Garabedian, M. J., Yamamoto, K. R. & Simons, S. S. J. (1991) *J. Biol. Chem.* **266**, 22075-22078). Furthermore, while ligand binding may suffice to trigger efficient cytoplasm-to-nuclear translocation, it may not necessarily cause binding of the receptor to its nuclear target and it may cause varying degrees of activation of the target gene.

Role of Intracellular Free Calcium and Energy in GFP-GR Translocation. Two important issues concerning the nuclear import of proteins were also addressed; these include the role of Ca^{++}, and the energy requirement of translocation. Intracellular stores of Ca^{++} were depleted by incubating the cells for 1 hour with the endoplasmic reticulum Ca^{++}-ATPase inhibitor, thapsigargin, and the calcium ionophore, ionomycin, in calcium-free media (intracellular free calcium content was measured with ratio imaging in Fura-2 loaded cells). The cytoplasmic pattern of GFP-GR was not significantly altered by calcium depletion. When Ca^{++}-depleted cells were subsequently exposed to dexamethasone (10nM for 30 min. at 37°C) in Ca^{++}-free media, the hormone induced complete translocation of GFP-GR, as seen by images taken from living cells with a confocal laser scanning microscope.

To study the energy dependence of ligand binding, cells were exposed to dexamethasone (10nM) at 4°C; then hormone was removed and the cells were warmed to 37°C under continuous monitoring with video-microscopy (cooled CCD camera from Axiovert 10 microscope system). At 4°C, translocation was completely arrested.

Rewarming led to complete translocation and reappearance of the focal GFP-GR localization. This experiment indicates that hormone binding to GR in living cells does not require energy in contrast to the energy-dependent step of translocation.

Focal Accumulation of Nuclear GFP-GR Correlates with Transcriptional Activation. When the intranuclear accumulation of GFP-GR is examined in detail, it is readily apparent that the receptor localizes most prominently at specific foci within the nucleus. In addition, there is a low level of accumulation in a diffuse reticular pattern, forming the basis for the nuclear background fluorescence. The number of these focal accumulations are unique to dexamethasone-treated cells and are not observed in 17-β-estradiol- or progesterone-exposed cells. In RU486-treated cells, focal points are not readily discernible. Instead, GFP-GR accumulates in a diffuse pattern with regions of condensation in a reticular pattern, such that regions of bright fluorescence appear thread-like in shape instead of as distinct foci. Depleting intracellular Ca^{++} did not affect the dexamethasone-mediated formation of intranuclear foci. The ability of agonist to induce focal accumulation of GFP-GR correlated strongly with its ability to activate transcription (Fig. 3C). The striking accumulation of dexamethasone-activated GFP-GR into intranuclear foci immediately suggests that a specific architecture may underlie this distribution. To further examine the structure of intranuclear GR binding sites, confocal laser scanning fluorescent microscopy and three dimensional image reconstruction was carried out.

Organized Architecture of Interphase Nuclei As Revealed by GFP-GR. Three-dimensional architecture of GFP-GR nuclear target sites was analyzed. Serial 0.5 μm sections of nuclei from dexamethasone treated cells were collected with confocal laser scanning fluorescent microscope, digitized images were imported into a Silicon Graphics Indigo 2 workstation, and three dimensional image segmentation, rendering, analysis and reconstruction was carried out with the ANALYZE software. GFP-GR distributions in the nuclei are displayed as pseudocolored, voxel-gradient-shaded, three dimensional projections. Three dimensional image analysis of the points of GR accumulation in dexamethasone-treated cells reveals a non-random distribution of

GFP-GR accumulation. Most strikingly, comparison of adjacent cells demonstrates a reproducible pattern of intranuclear structure for GFP-GR accumulation. A predominance of GFP-GR-accumulating foci is observed in the quadrant of the nucleus adjacent to the glass attachment surface of the cell, while a group of large patches of GFP-GR-containing foci are observed in the top half. Nucleolar structures were always devoid of GFP-GR. The nuclear pattern of RU486-treated cells was again strikingly different from dexamethasone-treated cells. Although essentially all of the GFP-GR is translocated, intranuclear RU486-liganded receptor is distributed throughout the nucleus in a reticular pattern but excluding nucleoli.

3134 cell line and derivatives

The 3134 cell line was derived from a mouse line designated 904.1. This cell was established by transfection of a murine mammary carcinoma line (C127) with a plasmid containing three functional segments: a) the bovine papilloma virus (BPV) 69% transforming fragment serves as a replicon in mammalian cells, b) the mouse mammary tumor virus (MMTV) LTR is a steroid responsive promoter and contains the GR binding sites, and c) the Ha-v-ras gene is a transforming oncogene and serves as a reporter for the MMTV promoter.

This plasmid replicates in 904.1 cells as an 9 kb episomal circle. During passage of this cell line, a spontaneous integration event occurred. This event resulted in the integration of a tandem array of the BPV/MMTV-LTR/Ha-v-ras cassette in perfect head-to-tail orientation. The integrated structure is diagrammed in Figure 2. Standard agarose gel electrophoresis and southern transfer hybridization analysis with a BPV probe (standard gel) was performed. CHEF gel high molecular weight analysis, again with a BPV probe (CHEF gel) was also performed. When the integrated array is digested with a one-cut restriction enzyme, the repeat unit of 8.8 kd is liberated as a single fragment (BamHI). NdeI and EcoRV (both no-cutters for the repeat unit) digestion leads to no digestion in the standard gel; EcoRV liberates a fragment approximately 2.2 megabase pairs in size. Enzymes that cut multiple times (PstI) give rise to the appropriate fragments for the repeat unit. Since the one cutter enzyme

releases only one size fragment from the array, the units must be organized in a perfect head-to-tail array.

Derivatives of 3134 cell lines containing pTET-nGFP-C656G and pOP-nGFP-C656G have been generated allowing GFP-GR expression upon tetracycline withdrawal from 5 ug/ml or upon induction with IPTG using standard procedures and as recommended by the manufacturer of these inducible systems, respectively. For example, 3616 is a single cell clone of 3134 with the pTET-nGFP-C656G DNA allowing acceptable level of GFP-GR expression that contains 200 copies of the MMTV LTR-cat-BPV tandem repeats; 3617 is another single cell clone of 3134 with the pTET-nGFP-C656G DNA but where the copy number of the MMTV LTR-cat-BPV tandem repeats dropped from 200 copies to 150 copies upon passage and then increased to about 170 copies in all cells upon further passage; 3596 is a single-cell clone of 3134 having 200 copies of the MMTV LTR-cat-BPV repeats and contains pOP-nGFP-C656G DNA; 3597 is a single cell clone of 3134 with 200 copies of the MMTV LTR-cat-BPV repeats and contains pOP-nGFP-C656G DNA. Acceptable levels of GFP-GR is reached in these derivatives after overnight induction of the regulatable promoters, e.g., after withdrawal of tetracycline in the case of 3616 and 3617, and after addition of IPTG in the case of 3596 and 3597.

Gene Targeting with GFP-GR in 3134 Cell

When the GFP-Glucocorticoid Receptor is transfected into the 3134 cell line, it is initially localized completely in the cytoplasm of the cell, as is normal, non-derivatized glucocorticoid receptor. When GFP-GR is activated with the GR ligand dexamethasone, the receptor translocates to the nucleus and accumulates on the BVP/MMTV-LTR/ras tandem array.

Figure 2 is a schematic representation of the appearance of GFP-GR when the fluorescent tag is excited by 489 nm laser light, and the 511 nm emission examined by confocal microscopy, using a standard fluorescein filter set. Individual sections are depicted for a representative 3134 nucleus. As one focuses on 0.18 micrometer sections

through the nucleus, a continuous fiber of intense light emission is detected over 5-7 sections, which corresponds to GFP-GR binding to the continuous BVP/MMTV-LTR/ras array.

Expression of GFP-GR in 3134 Cell

1) Materials.

a) Recombinant DNA

The LacSwitchTM inducible mammalian expression system (Stratagene catalog number 217450) is supplied with the phagemid DNAs: p3'SS for constitutive expression of the E. coli lac repressor and hygromycin resistance drug selectable marker; pOPI3 CAT for lac repressor regulated expression from the Rous sarcoma virus (RSV)-LTR promoter and neomycin resistance drug selectable marker; and pOPRSVI CAT for lac repressor regulated expression from the Rous sarcoma virus (RSV)-LTR promoter and neomycin resistance drug selectable marker.

Using standard recombinant DNA methodology, cDNA encoding GFP-GR fusion (either with the C656G mutation in the steroid binding domain of rat GR as present in the plasmid pCI-nGFP-C656G or wild type ligand binding domain of rat GR as present in the plasmid pCI-nGFP-rGR) is subcloned into the vector pOPRSVI CAT phagemid (Stratagene) at the NotI site. The cDNA should contain at its 5' end Kozak consensus sequence for efficient utilization of the translational initiation codon ATG, and at its 3' end the translational termination sequence followed by multiple translation termination sequences in all three reading frames. The vector provides RSV-LTR that has been engineered to be regulated by E. coli lac repressor so that in the mammalian cells with lac repressor, the strong RSV LTR promoter is inducible by the addition of 5 mM isopropyl-beta-D-thiogalactopyranoside (IPTG). The vector also provides an intron in the 5' untranslated region to ensure proper processing of the nascent transcript and maturation into productive mRNA. The vector also provides at the 3' untranslated region, herpes simplex thymidine kinase (TK) polyadenylation signal to ensure polyadenylation and message stability.

All DNAs used for introduction into mammalian cells are prepared from E. coli by alkali lysis procedure followed by isopycnic centrifugation by banding twice in cesium chloride/ethidium bromide gradient. Ethidium is removed by repeated extraction with isopentyl alcohol and cesium chloride by dialysis against large volumes of 10 mM Tris-HCl, pH 8/1mM EDTA (TE). After dialysis, DNA is extracted twice in phenol chloroform solution and then with chloroform before precipitation with 0.2 M sodium acetate (pH 5.5) and 70% ethanol at -20C. After collecting the precipitate by centrifugation, the DNA is washed with cold 70% ethanol, air dried, and then resuspended at a concentration of 1-2.5 mg/ml in TE.

b) Cell Line

The 3134 cell line contains a 9 kb repeat of the MMTV LTR fused to Ha-v-ras protooncogene in a BPV-based mammalian vector transformed into mouse carcinoma cell line C127.

Derivatives of 3134 cell lines containing pTET-nGFP-C656G and pOP-nGFP-C656G have been generated allowing GFP-GR expression upon tetracycline withdrawal from 5 ug/ml or upon induction with IPTG (about 1mM) using standard procedures and as recommended by the manufacturer of these inducible systems. Acceptable levels of GFP-GR is reached in these derivative cells after overnight induction of the regulatable promoters.

2) Maintenance of 3134 Cells

The 3134 cells are maintained in complete DMEM media [Dulbecco's Modified Eagle Medium (DMEM; Gibco-BRL catalog number 11965-084) supplement with 2 mM L-glutamine (Gibco-BRL catalog number 25030-024), 50ug/ml gentamicin reagent (Gibco-BRL catalog number 15750-011), and 10% fetal bovine serum (Gibco-BRL catalog number 26140-079) as monolayer in 162 cm^2 cell culture flasks (Costar catalog number 3150) at 37C in 5% CO_2 humidified air incubator. Upon confluence, cells are washed with Dulbecco's phosphate buffered saline (D-PBS) without calcium or magnesium and then treated with 6 ml of 0.05% trypsin/0.53 mM EDTA for several

minutes at RT (Gibco-BRL catalog number 25300-062). After cells round up, the side of the flask is tapped to dislodge cells from the flask surface. Repeated pipetting of the cells about three times results in a single cell suspension. One fifth of the cell suspension is transferred to a fresh 162 cm^2 flask containing 25 ml of the complete DMEM media. The freshly diluted cells are returned to the 37C/5% CO2 incubator.

3) Introduction of DNA into 3134 Cells by Electroporation

3134 cells grown to about 90% confluence are harvested using the trypsin-EDTA and the activity of the trypsin quenched by placing the single cell suspension into at least equal volume of complete DMEM. Cells are counted in a hemacytometer and concentrated by centrifugation in bench top centrifuge (Sorvall RT6000D at 2,000 rpm for 5 minutes at 4C) after placing in 50 ml sterile conical tube (Falcon 2070). Supernatant is aspirated and cells are washed again using complete DMEM. After the second wash, the resulting cell pellet is resuspended at a concentration of 20 million cells per 200 ul in cold DMEM and placed on ice. In 1.7 ml sterile microtube, appropriate DNA to be transfected, from 1-50ug of about 8 kb plasmid DNA (typically 1-20 ug), is placed. 200 ul 3134 cells is added and mixed well. The mixture is transferred to disposable electro chamber cuvettes (Gibco-BRL catalog number 11601-010) on ice. Cuvettes are placed into electroporation chamber and electroporated (Gibco-BRL Cell-Porator catalog series 1600; settings: 250 volts, 800 or 1180 microFarads, low resistance). Recovery on ice is allowed for 5 minutes. Then, cells are removed from the cuvette and transferred to 50 ml conical tube with complete DMEM at RT. Cells are plated out on 150 x 25 mm tissue culture dish (Falcon 3025) in 30 ml of complete DMEM containing about one fourth the content of the electroporation cuvette. Plates are placed in 37C/5% CO2 incubator. Next day, media is changed to fresh media and selected with appropriate drug. Every two days media is changed and selection maintained with the appropriate drug.

4) Stable 3134 Cell Line Containing Lac Repressor

Electroporation of 1 ug, 5 ug, 10 ug, 30 ug, and 60 ug p3'SS DNA into 20 million 3134 cells in 200 ul cold DMEM is performed as described above. One day

following electroporation, cells are selected with 450 ug/ml hygromycin B (Calbiochem catalog number 40005) for about 10 days during which hygromycin resistant colonies arise. The colonies are isolated by treating with trypsin impregnated sterile 3 mm cloning discs (PGC Scientifics catalog number 09-060) and transferred into 24 well tissue culture cluster dish (Costar 3524; each well 16 mm) containing 2 ml of complete DMEM with 450ug/ml hygromycin B. The cells are allowed to grow in the dish in a 37C/5% CO2 incubator until at least 50% confluence. Media is replaced every couple of days.

Cells are trypsinized and transferred to 100 x 20 mm dish (Falcon 3003) containing 10 ml of complete DMEM with 450 ug/ml hygromycin B. Then, at about confluence, cell are trypsinized and transferred to 150 x 25 mm dish with 30 ml of complete DMEM with 450 ug/ml hygromycin. Near confluence, cells were trypsinized and aliquot replated into 100 mm dish while the remaining cells are frozen down in 10% DMSO with complete DMEM at -70C. Cells grown on 100 mm dishes are harvested at confluence. Extracts are prepared and tested by Western blotting with rabbit anti-lac repressor polyclonal antibody (Stratagene catalog number 217449) and detected using goat anti-rabbit IgG horseradish peroxidase conjugated antibody (Bio-rad catalog number 170-6515) and enhanced chemiluminescent ECL western blotting detection reagents (Amersham catalog number RPN 2106). Cells expressing high amounts of the lac repressor are recloned and used for electroporating the pOPRSVI CAT plasmid containing GFP-C656G or GFP-rGR fusion cDNA.

5) Stable 3134 Cell Line Containing IPTG-Inducible GFP-GR or GFP-Fusion Protein Expression System.

3134 cells showing high levels of constitutive lac repressor expression without cytotoxic effects are used for electroporating pOPRSVI CAT plasmid containing the appropriate GFP-fusion protein cDNA. Following similar regiment as described above, the cells are selected with 450 ug/ml hygromycin B as well as 500 ug/ml geneticin (Gibco-BRL catalog number 11811-031). Following about ten days after selection,

again drug resistant colonies are isolated, expanded, and characterized. Characterization is achieved by examining uninduced as well as 5 mM IPTG induced cells using Southern blot analysis of the genomic DNA, Western blot analysis using antibody directed against GFP and GR or other steroid receptor, characterize for localization of the GFP-fusion protein on the MMTV LTR-tandem array upon exposure to the appropriate ligand. Cells with appropriate desirable characteristics are subcloned and serve as reagents for diagnostic purposes.

Additional GFP-Steroid Fusion Proteins

The same technology is used to prepare GFP fusions for each member of the steroid/thyroid receptor family, including the estrogen receptor, progesterone receptor, androgen receptor, mineralocorticoid receptor, thyroid receptor, retinoic acid receptor (RAR), RXR receptor, vitamin D receptor, and TCCD (dioxin) receptor. Chimeric proteins for each of the receptors is be prepared using GFP fused either to the N-terminus or the C-terminus of the appropriate receptor. The chimeras are tested for functional transcriptional transactivation activity, and for binding to cognate response elements using cell lines described below.

GFP-ER: For example, the glucocorticoid receptor portion of pCI-nGL1-C656G DNA has been replaced with a human estrogen receptor. The GFP-ER was found to be functional in both transcriptional activation as well as proper subcellular localization. This further demonstrates the utility of the expression vector as well as the GFP-tagging approach.

Modification of the GFP fluorescent moiety

The success of GFP-GR fusion in detecting and characterizing *in vivo* gene targeting indicates that modification and enhancement of the GFP polypeptide can be of considerable usefulness. These modifications include:

1) Changes in the exicitation wavelength to permit activation of selected GFP fusions;
2) Changes in the emission wavelength to permit observation of selected GFP fusions (GFP's will emit with different colors);

3) Enhancements to the efficiency of excitation and emission (GFP's will be "brighter");

4) Enhancements of GFP-chimeric proteins production by engineering the GFP sequence for more efficient expression:

a) increasing the steady state level of GFP-chimeric protein mRNA levels through enhanced transcription, transcript processing, and RNA stability;

b) increasing translational efficiency of the GFP-chimeric proteins by conservative substitutions of the protein coding regions and/or by introducing codons preferred for translation in humans cells;

c) increasing stability of the GFP-chimeric proteins by identification of protein destabilizing sequences and selective elimination of such turnover signals;

5) Changes in GFP to define the minimal fluorescing polypeptide that could function in a fusion protein.

For example, in plasmids pCI-nGL2-C656G; pCI-nGL3-C656G; pCI-nGL4-C656G; pCI-nGL5-C656G; pCI-nGL7-C656G; pCI-nGL9-C656G; pCI-nGL10-C656G; pCI-nGL11-C656G, blue variants of GFP are used. Additionally, in these plasmids, humanized codons encoding the fusion protein were utilized.

Additional Cell Lines

Human and other mammalian cell lines can be prepared containing receptor binding sites (receptor response elements) in multimerized arrays for direct visualization of *in vivo* gene targeting.

1) For each receptor, these arrays can be created by gene amplification. The receptor response element is fused to the dihydrofolate reductase (dHFR) gene, transfected into human cells, and integrants selected by resistance to methotrexate. After selection of the initial cell lines, further selection is applied using sequentially higher levels of methotrexate. This results in amplification of the dHFR/receptor response element sequence, providing an array to which the receptor/GFP fusion binding can be observed (directly analogous to the 3134 cell line for GR). Further detatils are provided below.

2) An alternative approach is to multimerize the receptor response element by synthetic DNA synthesis, then introduce this amplified element directly into mammalian cells. This can provide a concentrated target for GFP-receptor localization.

Generating Cells having Tandem Arrays

An approach that is being used to generate tandem arrays in a re-engineered promoter to obtain a functional transcription unit:

1) the MMTV LTR in the plasmid pLTRLuc is being mutated by site directed mutagenesis to introduce appropriate restriction enzyme cleavage sites at approximately positions:

+110 from start of transcription (HindIII)

-40 from start of transcription (SalI)

-217 from start of transcription (XhoI)

-1100 from start of transcription (XmaI)

2) the MMTV LTR in the plasmid pLTRLuc is being mutated by site directed mutagenesis to introduce appropriate restriction enzyme cleavage sites at approximately positions:

+110 from start of transcription (HindIII)

-80 from start of transcription (SalI)

-217 from start of transcription (XhoI)

-1100 from start of transcription (XmaI)

3) the DNA from (1) and (2) are being subcloned into two luciferase reporter gene vectors from Promega called pGL3-Basic (catalog number 88-1737) and pRL-CMV (catalog number 1068-2003) by using HindIII and XmaI digests of (1) and (2) and pGL3-Basic DNAs and inserting the mutated LTR fragments from (1) and (2) into the polylinker region of pGL3-Basic. In the case of pRL-CMV, this DNA is cleaved with PstI and BglII and the mutated LTR fragments from (1) and (2) inserted in place of the CMV Immediate Early Enhancer/Promoter regions in the pRL-CMV vector using a PstI/HindIII adapter on one end of the fragment. Also, before the insertion of the mutated LTR DNAs from (1) and (2), the SalI site present in the pGL3-Basic is

eliminated by digesting this vector with SalI, filling-in the end with DNA polymerase, and subsequent resealing of the filled-in end before transforming bacteria.

4) oligonucleotides containing the glucocorticoid response elements (GRE: 5' AGAACAnnnTGTTCT 3') (SEQ ID NO:4) or estrogen response elements (ERE: 5' AGGTCAnnnTGACCT 3') (SEQ ID NO:5) are synthesized and then annealed (e.g., oligonucleotide 1 and 2; or oligonucleotide 3 and 4) such that one end contains a cohesive end for SalI and the other XhoI. Some examples of such oligonucleotides are (where the lower case letters are spacer bases and in the above designation would have been designated as "n"):

oligonucleotide 1: (SEQ ID NO:6):
5' tcgagcgcgcaAGAACAcagTGTTCTgacgacacgaAGAACAggaTGTTCTcgtacagtg 3'
oligonucleotide 2: (SEQ ID NO:7):
5' tcgacactgtacgAGAACAtccTGTTCTtcgtgtcgtcAGAACActgTGTTCTtgcgcgc 3'
oligonucleotide 3: (SEQ ID NO:8):
5' tcgagcgcgcaAGGTCAcagTGACCTgacgacacgaAGGTCAggaTGACCTcgtacagtg 3'
oligonucleotide 4: (SEQ ID NO:9):
5' tcgacactgtacgAGGTCAtccTGACCTtcgtgtcgtcAGGTCActgTGACCTtgcgcgc 3'

5) the annealed oligoncleotides (1 and 2; or 3 and 4) are then ligated to generate large arrays in a perfect head-to-tail tandem array; these arrays may go through subcloning steps in bacteria to build larger arrays from smaller ones as well as to verify the integrity of the DNA sequence in the array by DNA sequencing.

6) the multimerized arrays from (5) are then inserted into the SalI/XhoI site in the mutated LTR of (3).

7) also, an additional series of constructs are also made lacking the region from -217 to
-1100 of the LTR by restriction enzyme digestion with Xho I and MluI for the pGL3-Basic vector derived clones of (6), filling in the XhoI/MluI cohesive ends, and resealing the blunt ends to generate a multimerized tandem array of a DNA binding site driving the expression of the luciferase gene only through the binding sites in the array and the region of the LTR containing the signal for the start point of transcription (the TATA box and the initiator sequence).

8) stable mammalian cell lines (e.g., from human origin or Chinese hamster ovary (CHO))are generated with the DNAs from (6) and (7) after linearization of the DNA, e.g. with BamHI and KpnI for the pGL3-Basic based DNAs and BamHI for the pRL-CMV based DNAs, and transfecting or electroporating these DNAs into cells, as is commonly done to introduce foreign DNA into cell. Also included in the transfection or electroporation is a DNA for a drug selectable marker such as neomycin gene or hygromycin that allows for drug selection (e.g., G418 or hygromycin, respectively) of cells that have taken up the foreign DNA. In addition, green fluorescent protein (GFP)-tagged glucocorticoid, estrogen, and orphan receptors or DNA binding proteins appropriate for the factor DNA-binding DNA element is also co-transfected or electroporated.

9) isolated clones from drug selected cells from (8) are characterized for the number of copies and dispersal of the introduced DNAs from (6) and (7). Also, characterized is the amount of fluorescence derived from the GFP tagged-receptors or DNA-binding protein as well as the functionality of the expressed protein.

10) because most of the cell lines from (9) will have the foreign DNA integrated throughout the genome, localization of the GFP-tagged DNA binding molecule on the arrays in each of these transcription unit would result in intense fluorescence spot or focus. It is also possible that an extremely few number of cells will have integrated the foreign DNA in small arrays which should look like slight elongated intense spot. The desirable cells are those that contain visible intense foci due to binding of the GFP-tagged DNA binding protein on the transcription unit. In the case of the glucocorticoid receptor, the desirable cell lines are such that agonist dexamethasone treatment results in accumulation of the GFP-GR on the transcription unit to generate intense foci but not antagonist RU486 treatment.

11) DNAs from (6) or (7) will also be ligated to generate tandem arrays of each transcription unit and cloned into cosmid vector or phage P1 vector.

12) the tandem array transcription units will be treated as in (8).

13) the result of (11) will be characterized as in (9) and (10) to obtain cell lines with desirable properties.

14) DNA from (6) and (7) will be linked to a constitutively expressible dihydrofolate reductase (DHFR) gene by standard DNA subcloning techniques.

15) the DNA from (14) will be introduced into mammalian cells lacking the tumor suppresor gene, p53, using a similar procedure as in (8) except leaving out the drug selectable marker DNAs (due to the fact that DHFR gene is a drug selectable marker gene) and then selected for two weeks with methotrexate at a concentration about four times above the LD90. Selection media is replaced every 2-3 days. After methotrexate resistant colonies appear (in approximately two weeks), the colonies are pooled and part of the pool is frozen down while the remainder is used in a repeat selection with a concentration of methotrexate about four times higher than that previously used. This last selection procedure is repeated a number of times to finally select out cells with a highly amplified tandem array copies of DNAs containing the multimerized binding sites, the reporter gene, and the DHFR gene.

16) the cell lines after the last methotrexate selection are characterized for the presence of at least one tandem array and localization of the GFP-tagged DNA binding protein on this array analyzed. Localization on this array is manifested as a long linear highly fluorescent structure observed in the nucleus. In the case of GFP-GR, the localization on the array should occur upon treatment with agonist dexamethasone but not antagonist RU486.

17) other approaches to generating array envisioned is the use of site specific recombinases in vivo to generate precisely tailored tandem arrays.

18) while items 1-16 focuses on the use of a genetically engineered artificial promoter/enhancer to generate arrays, tandem arrays may be made from just binding sites only or from natural regulatory DNAs harboring desirable factor-DNA binding sites.

19) since the higher eukaryotic genomes contain naturally occuring repetitive sequences, interactions of appropriately tagged DNA-binding protein with such naturally occurring sequences may also prove useful.

Using this general approach, functional receptor/cell line pairs can be established for each receptor-GFP fusion. These reagents provide a simple, rapid, straightforward, sensitive, and biologically relevant assay for each receptor.

Using standard recombinant DNA methodology, DNA encoding GFP-fusion protein sequences can be placed under the control of high expression eukaryotic promoter/enhancers (e.g. CMV promoter/enhancer, SV40 promoter/enhancer, RSV LTR,herpes simplex thymidine kinase [TK] promoter, etc.), naturally occurring inducible promoter/enhancers (e.g. metallothionine promoter/enhancer, MMTV LTR, heatshock promoter/enhancer, etc.), or synthetic inducible promoter/enhancers (e.g. GAL4-VP16 inducible system, Stratagene's LacSwitchTM inducible mammalian expression system (catalog number 217450), Life Technologies' Tet regulated expression system (catalog number 10583-011), etc.). These DNAs are introduced into mammalian cells along with DNA expressing a selectable marker (neomycin, hygromycin, zeocin, etc.) or screenable marker (e.g. fluorescence, foci formation, etc.) by standard protocol (e.g.calcium phosphate co-precipitation, electroporation, liposome-mediated transfection, viral infection, etc.). Following recovery of the cells to the introduction of DNA, approximately 1-3 days later, selectable agent is applied in the case of drug selectable marker to select for cells with stable integration of the selectable marker.

Upon continuous culturing of the cells under selection condition for an additional week during which clonal populations of cells will arise, the resulting cells are then characterized for the presence of the stably integrated DNA by Southern blot analysis and PCR analysis, for the expression of the GFP-fusion protein by Western blot analysis, flow cytometry, and microscopic examination, for participation in ligand-dependent translocation/transactivation by assaying the activity of reporter genes (e.g.chloramphenicol acetyltransferase, luciferase, beta-galactosidase, etc.) under the control of the appropriate ligand inducible promoter/enhancer (e.g., glucocorticoid response element containing promoter, estrogen response element containing promoter, etc.), and for localization on the target sequences in nuclei of expressing cells by high

resolution fluorescence imaging systems (e.g. confocal laser scanning microscopy, cooled CCD camera microscopy, etc.). Alternatively, stable transformants expressing GFP-fusion proteins can be directly isolated by fluorescence activated cell sorting (FACS) using appropriate excitation wavelengths and emission detector.

Although less preferable at present, an alternative to isolating clones of stable transformants will be the isolation of pools of stable transformants. The use of cell line/expression system combination may necessitate such a generation of a pool of stable transformants.

Additionally, while it is presently most efficient to obtain stable cell lines, GFP-fusion expression plasmid can be transiently introduced into cells and analyzed for ligand- dependent translocation/nuclear targeting.

Use of Reagents

Because the use of GFP as a tag involves fusing a rather large protein (27 kDa) to GR, it was examined whether any GR activity has been compromised by GFP. To this end, GFP was fused to a rat GR with the C656G point mutation. This allowed selectively activation of GFP-GR without activating the endogenous receptor, and, thus, to assess the activity of the chimeric receptor independently of the endogenous GR. From the dose response curve and transcriptional activation of the MMTV LTR reporter gene (Fig. 3B), it is clear that GFP-GR can be selectively activated without activating the endogenous receptor.

By a number of criteria, GFP-GR functions very much like GR. In particular, the tagged receptor resides in the cytoplasm until activated by a ligand; it then translocates into the nucleus at a rate comparable to that previously reported (22). The rate and extent of GFP-GR translocation shows a dependence on the concentration of the activating ligand as well as a ligand specificity reflective of the native receptor. Furthermore, since both dexamethasone and RU486 treatment lead to complete translocation of GFP-GR from the cyoplasm to the nucleus in all cells, essentially all of

the GFP-GR molecules exist in a conformation competent for both ligand binding and nuclear translocation. Once in the nucleus, GFP-GR's ability to activate the transcription of a MMTV LTR reporter gene depends on the type of activating ligand, consistent with previous results for GR. In the case of a potent agonist, dexamethasone, less GFP-GR is required for activation of transiently introduced MMTV LTRLuc reporter gene than for the endogenous GR, indicating that even with respect to transactivation potential, the presence of GFP has not altered the transcriptional potency ascribed to the C656G point mutation. Thus, in all aspects, the presence of GFP appears not to have affected normal GR function and has a general utility for studying the mechanisms of transcriptional regulation *in vivo*.

Since the S65T variant of GFP used here is highly excitable at 489 nm wavelength and resistant to photobleaching, it was possible to follow the course of cytoplasm-to-nuclear translocation of GR in a single living cell for an extended period of time. Upon binding to dexamethasone, GFP-GR moves vectorially toward the nucleus. Inhibition of import by chilling indicates that this transport is facilitated. Accumulation of GFP-GR along fibrillar structures before dexamethasone addition, and in the perinuclear region after dexamethasone addition, suggests that the cytoskeleton is involved in the transport process. Finally, the pulsatile brightness changes in the perinuclear region support an energy- and microtubule-dependent active transport process.

The rate of translocation of GFP-GR was dependent on hormone concentration, reflecting the dose- and time-dependence of GR action. This suggests that the rate of translocation contributes to GR function. The present invention indicates that cells having multiple copies of the response element in fluorescently detectable array can be a useful model to study reagents that modify rates of nuclear translocation and response element binding.

Finally, it is well-known that DNA in the nucleus is localized in a non-random fashion. The study of the pattern of nuclear fluorescence with GFP-GR supports the

idea of an inherent order in the organization of the interphase nuclei and may reflect structures related to this organization. The reproducibility in the pattern of GFP-GR accumulation between neighboring nuclei reflects an inherent order of the interphase nucleus with regards to both GR-target site architecture (the foci of bright fluorescence) as well as transcriptionally incompetent GFP-GR, which appears to accumulate in a reticular pattern, reminiscent of association with the nuclear matrix. Thus, functional differences in GFP-GR due to hormone-specific effects (e.g., dexamethasone vs. RU486) are reflected in the patterns of GFP-GR intranuclear accumulation, indicating a general utility of understanding intranuclear localization of GR in addressing hormone-mediated actions.

Throughout this application, various publications are referenced. The disclosures of these publications in their entireties are hereby incorporated by reference into this application in order to more fully describe the state of the art to which this invention pertains.

Although the present process has been described with reference to specific details of certain embodiments thereof, it is not intended that such details should be regarded as limitations upon the scope of the invention except as and to the extent that they are included in the accompaying claims.

Sequence Listing

(1) GENERAL INFORMATION:

(i) APPLICANT: The Government of the United States of America
as respresented by the Secretary
Department of Health and Human Services
Washington, D.C.

Htun Ph.D., Han
Hager Ph.D., Gordon L.

(ii) TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MONITORING DNA BINDING MOLECULES IN LIVING CELLS

(iii) NUMBER OF SEQUENCES: 9

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Needle & Rosenberg
(B) STREET: 127 Peachtree Street, Suite 1200
(C) CITY: Atlanta
(D) STATE: Georgia
(E) COUNTRY: USA
(F) ZIP: 30303

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: U.S. Serial No.60/008,373
(B) FILING DATE: 08 Dec 1995
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Selby, Elizabeth
(B) REGISTRATION NUMBER: 38298
(C) REFERENCE/DOCKET NUMBER: 14014.0183

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 404-688-0770
(B) TELEFAX: 404-688-9880

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7257 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(ii) MOLECULE TYPE: nucleic acid

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1072..4284

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
TCAATATTGG CCATTAGCCA TATTATTCAT TGGTTATATA GCATAAATCA ATATTGGCTA      60

TTGGCCATTG CATACGTTGT ATCTATATCA TAATATGTAC ATTTATATTG GCTCATGTCC     120

AATATGACCG CCATGTTGGC ATTGATTATT GACTAGTTAT TAATAGTAAT CAATTACGGG     180

GTCATTAGTT CATAGCCCAT ATATGGAGTT CCGCGTTACA TAACTTACGG TAAATGGCCC     240

GCCTGGCTGA CCGCCCAACG ACCCCCGCCC ATTGACGTCA ATAATGACGT ATGTTCCCAT     300

AGTAACGCCA ATAGGGACTT TCCATTGACG TCAATGGGTG GAGTATTTAC GGTAAACTGC     360

CCACTTGGCA GTACATCAAG TGTATCATAT GCCAAGTCCG CCCCCTATTG ACGTCAATGA     420

CGGTAAATGG CCCGCCTGGC ATTATGCCCA GTACATGACC TTACGGGACT TTCCTACTTG     480

GCAGTACATC TACGTATTAG TCATCGCTAT TACCATGGTG ATGCGGTTTT GGCAGTACAC     540

CAATGGGCGT GGATAGCGGT TTGACTCACG GGGATTTCCA AGTCTCCACC CCATTGACGT     600

CAATGGGAGT TTGTTTTGGC ACCAAAATCA ACGGGACTTT CCAAAATGTC GTAATAACCC     660

CGCCCCGTTG ACGCAAATGG GCGGTAGGCG TGTACGGTGG GAGGTCTATA TAAGCAGAGC     720

TCGTTTAGTG AACCGTCAGA TCACTAGAAG CTTTATTGCG GTAGTTTATC ACAGTTAAAT     780

TGCTAACGCA GTCAGTGCTT CTGACACAAC AGTCTCGAAC TTAAGCTGCA GAAGTTGGTC     840

GTGAGGCACT GGGCAGGTAA GTATCAAGGT TACAAGACAG GTTTAAGGAG ACCAATAGAA     900

ACTGGGCTTG TCGAGACAGA GAAGACTCTT GCGTTTCTGA TAGGCACCTA TTGGTCTTAC     960

TGACATCCAC TTTGCCTTTC TCTCCACAGG TGTCCACTCC CAGTTCAATT ACAGCTCTTA    1020

AGGCTAGAGT ACTTAATACG ACTCACTATA GGCTAGCGAA GGAGATCCGC C ATG GCC      1077
                                                          Met Ala
                                                            1

CAC CAT CAC CAC CAT CAC GGA TAT CCA TAC GAC GTG CCA GAT TAC GCT      1125
His His His His His His Gly Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
  5                  10                  15

CAG TCG AGT GCC ATG AGT AAA GGA GAA GAA CTT TTC ACT GGA GTT GTC      1173
Gln Ser Ser Ala Met Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val
 20                  25                  30

CCA ATT CTT GTT GAA TTA GAT GGT GAT GTT AAT GGG CAC AAA TTT TCT      1221
Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser
 35                  40                  45                  50

GTC AGT GGA GAG GGT GAA GGT GAT GCA ACA TAC GGA AAA CTT ACC CTT      1269
Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu
             55                  60                  65
```

```
AAA TTT ATT TGC ACT ACT GGA AAA CTA CCT GTT CCT TGG CCA ACA CTT      1317
Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu
          70                  75                  80

GTC ACT ACT TTC ACT TAT GGT GTT CAA TGC TTT TCA AGA TAC CCA GAT      1365
Val Thr Thr Phe Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp
      85                  90                  95

CAT ATG AAA CAG CAT GAC TTT TTC AAG AGT GCC ATG CCC GAA GGT TAT      1413
His Met Lys Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr
  100                 105                 110

GTA CAG GAA AGA ACT ATA TTT TTC AAA GAT GAC GGG AAC TAC AAG ACA      1461
Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr
115                 120                 125                 130

CGT GCT GAA GTC AAG TTT GAA GGT GAT ACC CTT GTT AAT AGA ATC GAG      1509
Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu
          135                 140                 145

TTA AAA GGT ATT GAT TTT AAA GAA GAT GGA AAC ATT CTT GGA CAC AAA      1557
Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys
        150                 155                 160

TTG GAA TAC AAC TAT AAC TCA CAC AAT GTA TAC ATC ATG GCA GAC AAA      1605
Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys
    165                 170                 175

CAA AAG AAT GGA ATC AAA GTT AAC TTC AAA ATT AGA CAC AAC ATT GAA      1653
Gln Lys Asn Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu
180                 185                 190

GAT GGA AGC GTT CAA CTA GCA GAC CAT TAT CAA CAA AAT ACT CCA ATT      1701
Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile
195                 200                 205                 210

GGC GAT GGC CCT GTC CTT TTA CCA GAC AAC CAT TAC CTG TCC ACA CAA      1749
Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln
          215                 220                 225

TCT GCC CTT TCG AAA GAT CCC AAC GAA AAG AGA GAC CAC ATG GTC CTT      1797
Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu
        230                 235                 240

CTT GAG TTT GTA ACA GCT GCT GGG ATT ACA CAT GGC ATG GAT GAA CTA      1845
Leu Glu Phe Val Thr Ala Ala Gly Ile Thr His Gly Met Asp Glu Leu
    245                 250                 255

TAC AAA GGC GCC GGC GCT GGT GCT GGT GCT GGC GCC ATC AGC GCG CTG      1893
Tyr Lys Gly Ala Gly Ala Gly Ala Gly Ala Gly Ala Ile Ser Ala Leu
260                 265                 270

ATC CTG GAC TCC AAA GAA TCC TTA GCT CCC CCT GGT AGA GAC GAA GTC      1941
Ile Leu Asp Ser Lys Glu Ser Leu Ala Pro Pro Gly Arg Asp Glu Val
275                 280                 285                 290

CCT GGC AGT TTG CTT GGC CAG GGG AGG GGG AGC GTA ATG GAC TTT TAT      1989
Pro Gly Ser Leu Leu Gly Gln Gly Arg Gly Ser Val Met Asp Phe Tyr
          295                 300                 305
```

```
AAA AGC CTG AGG GGA GGA GCT ACA GTC AAG GTT TCT GCA TCT TCG CCC      2037
Lys Ser Leu Arg Gly Gly Ala Thr Val Lys Val Ser Ala Ser Ser Pro
          310                 315                 320

TCA GTG GCT GCT GCT TCT CAG GCA GAT TCC AAG CAG CAG AGG ATT CTC      2085

Ser Val Ala Ala Ala Ser Gln Ala Asp Ser Lys Gln Gln Arg Ile Leu
      325                 330                 335

CTT GAT TTC TCG AAA GGC TCC ACA AGC AAT GTG CAG CAG CGA CAG CAG      2133
Leu Asp Phe Ser Lys Gly Ser Thr Ser Asn Val Gln Gln Arg Gln Gln
    340                 345                 350

CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG      2181
Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln
355                 360                 365                 370

CAG CCA GGC TTA TCC AAA GCC GTT TCA CTG TCC ATG GGG CTG TAT ATG      2229
Gln Pro Gly Leu Ser Lys Ala Val Ser Leu Ser Met Gly Leu Tyr Met
            375                 380                 385

GGA GAG ACA GAA ACA AAA GTG ATG GGG AAT GAC TTG GGC TAC CCA CAG      2277
Gly Glu Thr Glu Thr Lys Val Met Gly Asn Asp Leu Gly Tyr Pro Gln
          390                 395                 400

CAG GGC CAA CTT GGC CTT TCC TCT GGG GAA ACA GAC TTT CGG CTT CTG      2325
Gln Gly Gln Leu Gly Leu Ser Ser Gly Glu Thr Asp Phe Arg Leu Leu
      405                 410                 415

GAA GAA AGC ATT GCA AAC CTC AAT AGG TCG ACC AGC GTT CCA GAG AAC      2373
Glu Glu Ser Ile Ala Asn Leu Asn Arg Ser Thr Ser Val Pro Glu Asn
    420                 425                 430

CCC AAG AGT TCA ACG TCT GCA ACT GGG TGT GCT ACC CCG ACA GAG AAG      2421
Pro Lys Ser Ser Thr Ser Ala Thr Gly Cys Ala Thr Pro Thr Glu Lys
435                 440                 445                 450

GAG TTT CCC AAA ACT CAC TCG GAT GCA TCT TCA GAA CAG CAA AAT CGA      2469
Glu Phe Pro Lys Thr His Ser Asp Ala Ser Ser Glu Gln Gln Asn Arg
            455                 460                 465

AAA AGC CAG ACC GGC ACC AAC GGA GGC AGT GTG AAA TTG TAT CCC ACA      2517
Lys Ser Gln Thr Gly Thr Asn Gly Gly Ser Val Lys Leu Tyr Pro Thr
          470                 475                 480

GAC CAA AGC ACC TTT GAC CTC TTG AAG GAT TTG GAG TTT TCC GCT GGG      2565
Asp Gln Ser Thr Phe Asp Leu Leu Lys Asp Leu Glu Phe Ser Ala Gly
      485                 490                 495

TCC CCA AGT AAA GAC ACA AAC GAG AGT CCC TGG AGA TCA GAT CTG TTG      2613
Ser Pro Ser Lys Asp Thr Asn Glu Ser Pro Trp Arg Ser Asp Leu Leu
    500                 505                 510

ATA GAT GAA AAC TTG CTT TCT CCT TTG GCG GGA GAA GAT GAT CCA TTC      2661
Ile Asp Glu Asn Leu Leu Ser Pro Leu Ala Gly Glu Asp Asp Pro Phe
515                 520                 525                 530

CTT CTC GAA GGG AAC ACG AAT GAG GAT TGT AAG CCT CTT ATT TTA CCG      2709
Leu Leu Glu Gly Asn Thr Asn Glu Asp Cys Lys Pro Leu Ile Leu Pro
            535                 540                 545
```

```
GAC ACT AAA CCT AAA ATT AAG GAT ACT GGA GAT ACA ATC TTA TCA AGT      2757
Asp Thr Lys Pro Lys Ile Lys Asp Thr Gly Asp Thr Ile Leu Ser Ser
          550                 555                 560

CCC AGC AGT GTG GCA CTA CCC CAA GTG AAA ACA GAA AAA GAT GAT TTC      2805
Pro Ser Ser Val Ala Leu Pro Gln Val Lys Thr Glu Lys Asp Asp Phe
      565                 570                 575

ATT GAA CTT TGC ACC CCC GGG GTA ATT AAG CAA GAG AAA CTG GGC CCA      2853
Ile Glu Leu Cys Thr Pro Gly Val Ile Lys Gln Glu Lys Leu Gly Pro
    580                 585                 590

GTT TAT TGT CAG GCA AGC TTT TCT GGG ACA AAT ATA ATT GGT AAT AAA      2901
Val Tyr Cys Gln Ala Ser Phe Ser Gly Thr Asn Ile Ile Gly Asn Lys
595                 600                 605                 610

ATG TCT GCC ATT TCT GTT CAT GGT GTG AGT ACC TCT GGA GGA CAG ATG      2949
Met Ser Ala Ile Ser Val His Gly Val Ser Thr Ser Gly Gly Gln Met
            615                 620                 625

TAC CAC TAT GAC ATG AAT ACA GCA TCC CTT TCT CAG CAG CAG GAT CAG      2997
Tyr His Tyr Asp Met Asn Thr Ala Ser Leu Ser Gln Gln Gln Asp Gln
          630                 635                 640

AAG CCT GTT TTT AAT GTC ATT CCA CCA ATT CCT GTT GGT TCT GAA AAC      3045
Lys Pro Val Phe Asn Val Ile Pro Pro Ile Pro Val Gly Ser Glu Asn
      645                 650                 655

TGG AAT AGG TGC CAA GGC TCC GGA GAG GAC AGC CTG ACT TCC TTG GGG      3093
Trp Asn Arg Cys Gln Gly Ser Gly Glu Asp Ser Leu Thr Ser Leu Gly
    660                 665                 670

GCT CTG AAC TTC CCA GGC CGG TCA GTG TTT TCT AAT GGG TAC TCA AGC      3141
Ala Leu Asn Phe Pro Gly Arg Ser Val Phe Ser Asn Gly Tyr Ser Ser
675                 680                 685                 690

CCT GGA ATG AGA CCA GAT GTA AGC TCT CCT CCA TCC AGC TCG TCA GCA      3189
Pro Gly Met Arg Pro Asp Val Ser Ser Pro Pro Ser Ser Ser Ser Ala
            695                 700                 705

GCC ACG GGA CCA CCT CCC AAG CTC TGC CTG GTG TGC TCC GAT GAA GCT      3237
Ala Thr Gly Pro Pro Pro Lys Leu Cys Leu Val Cys Ser Asp Glu Ala
          710                 715                 720

TCA GGA TGT CAT TAC GGG GTG CTG ACA TGT GGA AGC TGC AAA GTA TTC      3285
Ser Gly Cys His Tyr Gly Val Leu Thr Cys Gly Ser Cys Lys Val Phe
      725                 730                 735

TTT AAA AGA GCA GTG GAA GGA CAG CAC AAT TAC CTT TGT GCT GGA AGA      3333
Phe Lys Arg Ala Val Glu Gly Gln His Asn Tyr Leu Cys Ala Gly Arg
    740                 745                 750

AAC GAT TGC ATC ATT GAT AAA ATT CGA AGG AAA AAC TGC CCA GCA TGC      3381
Asn Asp Cys Ile Ile Asp Lys Ile Arg Arg Lys Asn Cys Pro Ala Cys
755                 760                 765                 770

CGC TAT CGG AAA TGT CTT CAG GCT GGA ATG AAC CTT GAA GCT CGA AAA      3429
Arg Tyr Arg Lys Cys Leu Gln Ala Gly Met Asn Leu Glu Ala Arg Lys
            775                 780                 785

ACA AAG AAA AAA ATC AAA GGG ATT CAG CAA GCC ACT GCA GGA GTC TCA      3477
```

```
Thr Lys Lys Lys Ile Lys Gly Ile Gln Gln Ala Thr Ala Gly Val Ser
          790                 795                 800

CAA GAC ACT TCG GAA AAT CCT AAC AAA ACA ATA GTT CCT GCA GCA TTA      3525
Gln Asp Thr Ser Glu Asn Pro Asn Lys Thr Ile Val Pro Ala Ala Leu
      805                 810                 815

CCA CAG CTC ACC CCT ACC TTG GTG TCA CTG CTG GAG GTG ATT GAA CCC      3573
Pro Gln Leu Thr Pro Thr Leu Val Ser Leu Leu Glu Val Ile Glu Pro
    820                 825                 830

GAG GTG TTG TAT GCA GGA TAT GAT AGC TCT GTT CCA GAT TCA GCA TGG      3621
Glu Val Leu Tyr Ala Gly Tyr Asp Ser Ser Val Pro Asp Ser Ala Trp
835                 840                 845                 850

AGA ATT ATG ACC ACA CTC AAC ATG TTA GGT GGG CGT CAA GTG ATT GCA      3669
Arg Ile Met Thr Thr Leu Asn Met Leu Gly Gly Arg Gln Val Ile Ala
            855                 860                 865

GCA GTG AAA TGG GCA AAG GCG ATA CTA GGC TTG AGA AAC TTA CAC CTC      3717
Ala Val Lys Trp Ala Lys Ala Ile Leu Gly Leu Arg Asn Leu His Leu
          870                 875                 880

GAT GAC CAA ATG ACC CTG CTA CAG TAC TCA TGG ATG TTT CTC ATG GCA      3765
Asp Asp Gln Met Thr Leu Leu Gln Tyr Ser Trp Met Phe Leu Met Ala
      885                 890                 895

TTT GCC TTG GGT TGG AGA TCA TAC AGA CAA TCA AGC GGA AAC CTG CTC      3813
Phe Ala Leu Gly Trp Arg Ser Tyr Arg Gln Ser Ser Gly Asn Leu Leu
    900                 905                 910

TGC TTT GCT CCT GAT CTG ATT ATT AAT GAG CAG AGA ATG TCT CTA CCC      3861
Cys Phe Ala Pro Asp Leu Ile Ile Asn Glu Gln Arg Met Ser Leu Pro
915                 920                 925                 930

GGC ATG TAT GAC CAA TGT AAA CAC ATG CTG TTT GTC TCC TCT GAA TTA      3909
Gly Met Tyr Asp Gln Cys Lys His Met Leu Phe Val Ser Ser Glu Leu
            935                 940                 945

CAA AGA TTG CAG GTA TCC TAT GAA GAG TAT CTC TGT ATG AAA ACC TTA      3957
Gln Arg Leu Gln Val Ser Tyr Glu Glu Tyr Leu Cys Met Lys Thr Leu
          950                 955                 960

CTG CTT CTC TCC TCA GTT CCT AAG GAA GGT CTG AAG AGC CAA GAG TTA      4005
Leu Leu Leu Ser Ser Val Pro Lys Glu Gly Leu Lys Ser Gln Glu Leu
      965                 970                 975

TTT GAT GAG ATT CGA ATG ACT TAT ATC AAA GAG CTA GGA AAA GCC ATC      4053
Phe Asp Glu Ile Arg Met Thr Tyr Ile Lys Glu Leu Gly Lys Ala Ile
    980                 985                 990

GTC AAA AGG GAA GGG AAC TCC AGT CAG AAC TGG CAA CGG TTT TAC CAA      4101
Val Lys Arg Glu Gly Asn Ser Ser Gln Asn Trp Gln Arg Phe Tyr Gln
995                 1000                1005                1010

CTG ACA AAG CTT CTG GAC TCC ATG CAT GAG GTG GTT GAG AAT CTC CTT      4149
Leu Thr Lys Leu Leu Asp Ser Met His Glu Val Val Glu Asn Leu Leu
            1015                1020                1025
```

```
ACC TAC TGC TTC CAG ACA TTT TTG GAT AAG ACC ATG AGT ATT GAA TTC      4197
Thr Tyr Cys Phe Gln Thr Phe Leu Asp Lys Thr Met Ser Ile Glu Phe
        1030                1035                1040

CCA GAG ATG TTA GCT GAA ATC ATC ACT AAT CAG ATA CCA AAA TAT TCA      4245
Pro Glu Met Leu Ala Glu Ile Ile Thr Asn Gln Ile Pro Lys Tyr Ser
     1045                1050                1055

AAT GGA AAT ATC AAA AAG CTT CTG TTT CAT CAA AAA TGA CTGCCTTACT      4294
Asn Gly Asn Ile Lys Lys Leu Leu Phe His Gln Lys  *
    1060                1065                1070

AAGAAAGGTT GCCTTAAAGA AAGTTGAATT TATAGTCTAG AGTCGACCCG GGCGGCCGCT    4354

TCGAGCAGAC ATGATAAGAT ACATTGATGA GTTTGGACAA ACCACAACTA GAATGCAGTG    4414

AAAAAAATGC TTTATTTGTG AAATTTGTGA TGCTATTGCT TTATTTGTAA CCATTATAAG    4474

CTGCAATAAA CAAGTTAACA ACAACAATTG CATTCATTTT ATGTTTCAGG TTCAGGGGGA    4534

GATGTGGGAG GTTTTTTAAA GCAAGTAAAA CCTCTACAAA TGTGGTAAAA TCGATAAGGA    4594

TCCGGGCTGG CGTAATAGCG AAGAGGCCCG CACCGATCGC CCTTCCCAAC AGTTGCGCAG    4654

CCTGAATGGC GAATGGACGC GCCCTGTAGC GGCGCATTAA GCGCGGCGGG TGTGGTGGTT    4714

ACGCGCAGCG TGACCGCTAC ACTTGCCAGC GCCCTAGCGC CCGCTCCTTT CGCTTTCTTC    4774

CCTTCCTTTC TCGCCACGTT CGCCGGCTTT CCCCGTCAAG CTCTAAATCG GGGGCTCCCT    4834

TTAGGGTTCC GATTTAGAGC TTTACGGCAC CTCGACCGCA AAAAACTTGA TTTGGGTGAT    4894

GGTTCACGTA GTGGGCCATC GCCCTGATAG ACGGTTTTTC GCCCTTTGAC GTTGGAGTCC    4954

ACGTTCTTTA ATAGTGGACT CTTGTTCCAA ACTGGAACAA CACTCAACCC TATCTCGGTC    5014

TATTCTTTTG ATTTATAAGG GATTTTGCCG ATTTCGGCCT ATTGGTTAAA AAATGAGCTG    5074

ATTTAACAAA TATTTAACGC GAATTTTAAC AAAATATTAA CGTTTACAAT TTCGCCTGAT    5134

GCGGTATTTT CTCCTTACGC ATCTGTGCGG TATTTCACAC CGCATATGGT GCACTCTCAG    5194

TACAATCTGC TCTGATGCCG CATAGTTAAG CCAGCCCCGA CACCCGCCAA CACCCGCTGA    5254

CGCGCCCTGA CGGGCTTGTC TGCTCCCGGC ATCCGCTTAC AGACAAGCTG TGACCGTCTC    5314

CGGGAGCTGC ATGTGTCAGA GGTTTTCACC GTCATCACCG AAACGCGCGA GACGAAAGGG    5374

CCTCGTGATA CGCCTATTTT TATAGGTTAA TGTCATGATA ATAATGGTTT CTTAGACGTC    5434

AGGTGGCACT TTTCGGGGAA ATGTGCGCGG AACCCCTATT TGTTTATTTT TCTAAATACA    5494

TTCAAATATG TATCCGCTCA TGAGACAATA ACCCTGATAA ATGCTTCAAT AATATTGAAA    5554

AAGGAAGAGT ATGAGTATTC AACATTTCCG TGTCGCCCTT ATTCCCTTTT TTGCGGCATT    5614

TTGCCTTCCT GTTTTTGCTC ACCCAGAAAC GCTGGTGAAA GTAAAAGATG CTGAAGATCA    5674

GTTGGGTGCA CGAGTGGGTT ACATCGAACT GGATCTCAAC AGCGGTAAGA TCCTTGAGAG    5734

TTTTCGCCCC GAAGAACGTT TTCCAATGAT GAGCACTTTT AAAGTTCTGC TATGTGGCGC    5794
```

GGTATTATCC CGTATTGACG CCGGGCAAGA GCAACTCGGT CGCCGCATAC ACTATTCTCA 5854

GAATGACTTG GTTGAGTACT CACCAGTCAC AGAAAAGCAT CTTACGGATG GCATGACAGT 5914

AAGAGAATTA TGCAGTGCTG CCATAACCAT GAGTGATAAC ACTGCGGCCA ACTTACTTCT 5974

GACAACGATC GGAGGACCGA AGGAGCTAAC CGCTTTTTTG CACAACATGG GGGATCATGT 6034

AACTCGCCTT GATCGTTGGG AACCGGAGCT GAATGAAGCC ATACCAAACG ACGAGCGTGA 6094

CACCACGATG CCTGTAGCAA TGGCAACAAC GTTGCGCAAA CTATTAACTG GCGAACTACT 6154

TACTCTAGCT TCCCGGCAAC AATTAATAGA CTGGATGGAG GCGGATAAAG TTGCAGGACC 6214

ACTTCTGCGC TCGGCCCTTC CGGCTGGCTG GTTTATTGCT GATAAATCTG GAGCCGGTGA 6274

GCGTGGGTCT CGCGGTATCA TTGCAGCACT GGGGCCAGAT GGTAAGCCCT CCCGTATCGT 6334

AGTTATCTAC ACGACGGGGA GTCAGGCAAC TATGGATGAA CGAAATAGAC AGATCGCTGA 6394

GATAGGTGCC TCACTGATTA AGCATTGGTA ACTGTCAGAC CAAGTTTACT CATATATACT 6454

TTAGATTGAT TTAAAACTTC ATTTTTAATT TAAAAGGATC TAGGTGAAGA TCCTTTTTGA 6514

TAATCTCATG ACCAAAATCC CTTAACGTGA GTTTTCGTTC CACTGAGCGT CAGACCCCGT 6574

AGAAAAGATC AAAGGATCTT CTTGAGATCC TTTTTTTCTG CGCGTAATCT GCTGCTTGCA 6634

AACAAAAAAA CCACCGCTAC CAGCGGTGGT TTGTTTGCCG GATCAAGAGC TACCAACTCT 6694

TTTTCCGAAG GTAACTGGCT TCAGCAGAGC GCAGATACCA AATACTGTCC TTCTAGTGTA 6754

GCCGTAGTTA GGCCACCACT TCAAGAACTC TGTAGCACCG CCTACATACC TCGCTCTGCT 6814

AATCCTGTTA CCAGTGGCTG CTGCCAGTGG CGATAAGTCG TGTCTTACCG GGTTGGACTC 6874

AAGACGATAG TTACCGGATA AGGCGCAGCG GTCGGGCTGA ACGGGGGGTT CGTGCACACA 6934

GCCCAGCTTG GAGCGAACGA CCTACACCGA ACTGAGATAC CTACAGCGTG AGCTATGAGA 6994

AAGCGCCACG CTTCCCGAAG GGAGAAAGGC GGACAGGTAT CCGGTAAGCG GCAGGGTCGG 7054

AACAGGAGAG CGCACGAGGG AGCTTCCAGG GGGAAACGCC TGGTATCTTT ATAGTCCTGT 7114

CGGGTTTCGC CACCTCTGAC TTGAGCGTCG ATTTTTGTGA TGCTCGTCAG GGGGGCGGAG 7174

CCTATGGAAA AACGCCAGCA ACGCGGCCTT TTTACGGTTC CTGGCCTTTT GCTGGCCTTT 7234

TGCTCACATG GCTCGACAGA TCT 7257

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1071 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Ala His His His His His His Gly Tyr Pro Tyr Asp Val Pro Asp
 1               5                  10                  15

Tyr Ala Gln Ser Ser Ala Met Ser Lys Gly Glu Glu Leu Phe Thr Gly
            20                  25                  30

Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys
        35                  40                  45

Phe Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu
    50                  55                  60

Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro
65                  70                  75                  80

Thr Leu Val Thr Thr Phe Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr
                85                  90                  95

Pro Asp His Met Lys Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu
            100                 105                 110

Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr
        115                 120                 125

Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg
    130                 135                 140

Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly
145                 150                 155                 160

His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala
                165                 170                 175

Asp Lys Gln Lys Asn Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn
            180                 185                 190

Ile Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr
        195                 200                 205

Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser
    210                 215                 220

Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met
225                 230                 235                 240

Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr His Gly Met Asp
                245                 250                 255

Glu Leu Tyr Lys Gly Ala Gly Ala Gly Ala Gly Ala Gly Ala Ile Ser
            260                 265                 270

Ala Leu Ile Leu Asp Ser Lys Glu Ser Leu Ala Pro Pro Gly Arg Asp
        275                 280                 285

Glu Val Pro Gly Ser Leu Leu Gly Gln Gly Arg Gly Ser Val Met Asp
    290                 295                 300

Phe Tyr Lys Ser Leu Arg Gly Gly Ala Thr Val Lys Val Ser Ala Ser
305                 310                 315                 320
```

```
Ser Pro Ser Val Ala Ala Ala Ser Gln Ala Asp Ser Lys Gln Gln Arg
            325                 330                 335

Ile Leu Leu Asp Phe Ser Lys Gly Ser Thr Ser Asn Val Gln Gln Arg
        340                 345                 350

Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln
    355                 360                 365

Gln Gln Gln Pro Gly Leu Ser Lys Ala Val Ser Leu Ser Met Gly Leu
370                 375                 380

Tyr Met Gly Glu Thr Glu Thr Lys Val Met Gly Asn Asp Leu Gly Tyr
385                 390                 395                 400

Pro Gln Gln Gly Gln Leu Gly Leu Ser Ser Gly Glu Thr Asp Phe Arg
            405                 410                 415

Leu Leu Glu Glu Ser Ile Ala Asn Leu Asn Arg Ser Thr Ser Val Pro
        420                 425                 430

Glu Asn Pro Lys Ser Ser Thr Ser Ala Thr Gly Cys Ala Thr Pro Thr
    435                 440                 445

Glu Lys Glu Phe Pro Lys Thr His Ser Asp Ala Ser Ser Glu Gln Gln
450                 455                 460

Asn Arg Lys Ser Gln Thr Gly Thr Asn Gly Gly Ser Val Lys Leu Tyr
465                 470                 475                 480

Pro Thr Asp Gln Ser Thr Phe Asp Leu Leu Lys Asp Leu Glu Phe Ser
            485                 490                 495

Ala Gly Ser Pro Ser Lys Asp Thr Asn Glu Ser Pro Trp Arg Ser Asp
        500                 505                 510

Leu Leu Ile Asp Glu Asn Leu Leu Ser Pro Leu Ala Gly Glu Asp Asp
    515                 520                 525

Pro Phe Leu Leu Glu Gly Asn Thr Asn Glu Asp Cys Lys Pro Leu Ile
530                 535                 540

Leu Pro Asp Thr Lys Pro Lys Ile Lys Asp Thr Gly Asp Thr Ile Leu
545                 550                 555                 560

Ser Ser Pro Ser Ser Val Ala Leu Pro Gln Val Lys Thr Glu Lys Asp
            565                 570                 575

Asp Phe Ile Glu Leu Cys Thr Pro Gly Val Ile Lys Gln Glu Lys Leu
        580                 585                 590

Gly Pro Val Tyr Cys Gln Ala Ser Phe Ser Gly Thr Asn Ile Ile Gly
    595                 600                 605

Asn Lys Met Ser Ala Ile Ser Val His Gly Val Ser Thr Ser Gly Gly
610                 615                 620

Gln Met Tyr His Tyr Asp Met Asn Thr Ala Ser Leu Ser Gln Gln Gln
625                 630                 635                 640
```

```
Asp Gln Lys Pro Val Phe Asn Val Ile Pro Pro Ile Pro Val Gly Ser
            645                 650                 655

Glu Asn Trp Asn Arg Cys Gln Gly Ser Gly Glu Asp Ser Leu Thr Ser
        660                 665                 670

Leu Gly Ala Leu Asn Phe Pro Gly Arg Ser Val Phe Ser Asn Gly Tyr
    675                 680                 685

Ser Ser Pro Gly Met Arg Pro Asp Val Ser Ser Pro Pro Ser Ser Ser
690                 695                 700

Ser Ala Ala Thr Gly Pro Pro Pro Lys Leu Cys Leu Val Cys Ser Asp
705                 710                 715                 720

Glu Ala Ser Gly Cys His Tyr Gly Val Leu Thr Cys Gly Ser Cys Lys
            725                 730                 735

Val Phe Phe Lys Arg Ala Val Glu Gly Gln His Asn Tyr Leu Cys Ala
        740                 745                 750

Gly Arg Asn Asp Cys Ile Ile Asp Lys Ile Arg Arg Lys Asn Cys Pro
    755                 760                 765

Ala Cys Arg Tyr Arg Lys Cys Leu Gln Ala Gly Met Asn Leu Glu Ala
770                 775                 780

Arg Lys Thr Lys Lys Lys Ile Lys Gly Ile Gln Gln Ala Thr Ala Gly
785                 790                 795                 800

Val Ser Gln Asp Thr Ser Glu Asn Pro Asn Lys Thr Ile Val Pro Ala
            805                 810                 815

Ala Leu Pro Gln Leu Thr Pro Thr Leu Val Ser Leu Leu Glu Val Ile
        820                 825                 830

Glu Pro Glu Val Leu Tyr Ala Gly Tyr Asp Ser Ser Val Pro Asp Ser
    835                 840                 845

Ala Trp Arg Ile Met Thr Thr Leu Asn Met Leu Gly Gly Arg Gln Val
850                 855                 860

Ile Ala Ala Val Lys Trp Ala Lys Ala Ile Leu Gly Leu Arg Asn Leu
865                 870                 875                 880

His Leu Asp Asp Gln Met Thr Leu Leu Gln Tyr Ser Trp Met Phe Leu
            885                 890                 895

Met Ala Phe Ala Leu Gly Trp Arg Ser Tyr Arg Gln Ser Ser Gly Asn
        900                 905                 910

Leu Leu Cys Phe Ala Pro Asp Leu Ile Ile Asn Glu Gln Arg Met Ser
    915                 920                 925

Leu Pro Gly Met Tyr Asp Gln Cys Lys His Met Leu Phe Val Ser Ser
930                 935                 940

Glu Leu Gln Arg Leu Gln Val Ser Tyr Glu Glu Tyr Leu Cys Met Lys
945                 950                 955                 960
```

```
Thr Leu Leu Leu Leu Ser Ser Val Pro Lys Glu Gly Leu Lys Ser Gln
            965                 970                 975

Glu Leu Phe Asp Glu Ile Arg Met Thr Tyr Ile Lys Glu Leu Gly Lys
          980                 985                 990

Ala Ile Val Lys Arg Glu Gly Asn Ser Ser Gln Asn Trp Gln Arg Phe
      995                 1000                1005

Tyr Gln Leu Thr Lys Leu Leu Asp Ser Met His Glu Val Val Glu Asn
    1010                1015                1020

Leu Leu Thr Tyr Cys Phe Gln Thr Phe Leu Asp Lys Thr Met Ser Ile
1025                1030                1035                1040

Glu Phe Pro Glu Met Leu Ala Glu Ile Ile Thr Asn Gln Ile Pro Lys
            1045                1050                1055

Tyr Ser Asn Gly Asn Ile Lys Lys Leu Leu Phe His Gln Lys  *
          1060                1065                1070
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 46 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GCGCGCTGAT CAGAATTCCT TTTAGGAATT CTGATCAGCG CGCTGA 46

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AGAACANNNT GTTCT 15

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AGGTCANNNT GACCT 15

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TCGAGCGCGC AAGAACACAG TGTTCTGACG ACACGAAGAA CAGGATGTTC TCGTACAGTG 60

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

TCGACACTGT ACGAGAACAT CCTGTTCTTC GTGTCGTCAG AACACTGTGT TCTTGCGCGC 60

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TCGAGCGCGC AAGGTCACAG TGACCTGACG ACACGAAGGT CAGGATGACC TCGTACAGTG 60

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TCGACACTGT ACGAGGTCAT CCTGACCTTC GTGTCGTCAG GTCACTGTGA CCTTGCGCGC 60

What is claimed is:

1. A mammalian cell having a plurality of steroid receptor response elements in an array such that the response element can be directly detected when bound by fluorescently labeled steroid receptor.

2. The cell of claim 1, wherein the response element is integrated into the genome of the cell.

3. The cell of claim 1, wherein the response element is present in the mouse mammary tumor virus long terminal repeat.

4. The cell of claim 3, wherein the cell is a cell of the cell line designated 3134 deposited with American Type Culture Collection under accession number CRL-11998 (ATCC).

5. The cell of claim 1, further comprising a nucleic acid encoding a chimeric protein wherein a fluorescent protein is fused to the steroid receptor.

6. The cell of claim 5, wherein the nucleic acid is integrated into the genome of the cell.

7. The cell of claim 5, wherein the fluorescent protein is green fluorescent protein.

8. The cell of claim 5, wherein the steroid receptor is glucocorticoid receptor.

9. The cell of claim 8, wherein the chimeric protein has the amino acid sequence set forth in SEQ ID NO: 2 .

10. A method of screening for a compound that binds to a selected nucleic acid comprising:

    a. contacting compound fluorescently labeled by a fluorescent protein with a cell having a plurality of copies of the nucleic acid in an array such that the nucleic

acid can be directly detected when bound by fluorescently labeled compound; and

b. directly detecting the location of fluorescence within the cell,

fluorescence aggregated at the site of the nucleic acid array indicating a compound that binds to the selected nucleic acid.

11. The method of claim 10, wherein the compound is fluorescently labeled with a green fluorescent protein.

12. The method of claim 10, wherein the selected nucleic acid is integrated into the genome of the cell.

13. A method of screening for a ligand that activates gene targeting of a steroid receptor in the nucleus of a mammalian cell comprising:

a. contacting the ligand with the cell of claim 5; and

b. directly detecting the location of fluorescence within the cell,

fluorescence aggregated at the site of the steroid receptor response element array in the nucleus indicating a ligand that activates the gene targeting of a steroid receptor in the nucleus of a mammalian cell.

14. The method of claim 13, wherein the fluorescent protein is green fluorescent protein.

15. The method of Claim 13, wherein the cell is a cell of the cell line designated 3134 deposited with American Type Culture Collection under accession number _ CRL-11998 (ATCC).

16. A method of screening for a ligand that activates the translocation of a steroid receptor to the nucleus in a mammalian cell comprising:

a. contacting the cell of claim 5 with the ligand; and

b. directly detecting the location of fluorescence within the cell,

a change in the relative fluorescence of the nucleus to the cytoplasm such as to increase the fluorescence of the nucleus indicating a ligand that activates the translocation of a steroid receptor to the nucleus in a mammalian cell.

17. The method of claim 16, wherein the fluorescent protein is green fluorescent protein.

18. The method of claim 16, wherein the mammalian cell is a cell of the cell line designated 3134 deposited with American Type Culture Collection under accession number CRL-11998 (ATCC).

19. A method of detecting in a biological sample the presence of an agonist of a steroid receptor comprising:

a. contacting the sample with the cell of claim 5 and

b. directly detecting the location of fluorescence within the cell,

the location of fluorescence aggregated at the site of the steroid receptor response element array in the nucleus indicating the presence of an agonist of the steroid receptor in the sample.

20. The method of claim 19, wherein the fluorescent protein is green fluorescent protein.

21. The method of claim 19, wherein the cell is a cell of the cell line designated 3134 deposited with American Type Culture Collection under accession number CRL-11998 (ATCC).

22. A method of detecting in a biological sample the presence of an antagonist of a steroid receptor comprising:

a. contacting the sample and an agonist of the steroid receptor with the cell of claim 5 ; and

b. directly detecting the location of fluorescence within the cell,

the absence of fluorescence substantially aggregated at the site of the steroid receptor response element array in the nucleus indicating the presence of an antagonist of the steroid receptor in the sample.

23. The method of claim 22, wherein the fluorescent protein is green fluorescent protein.

24. The method of claim 22, wherein the cell is a cell of the cell line designated 3134 deposited with American Type Culture Collection under accession number CRL-11998 (ATCC).

25. A method of monitoring the level of an agonist of a steroid receptor in a subject comprising:

a. periodically obtaining a biological sample from the subject,

b. contacting the sample with the cell of claim 5, and

c. directly detecting the location of fluorescence within the cell,

a decrease in fluorescence aggregated at the site of the steroid receptor response element in the nucleus in a later-obtained sample relative to an earlier-obtained sample indicating a decrease in level of the steroid agonist of the steroid receptor in the sample and an increase in fluorescence aggregated at the site of the steroid receptor response element in the nucleus in a later-obtained sample relative to an earlier-obtained sample indicating an increase in level of the steroid agonist of the steroid receptor in the sample.

26. The method of claim 25, wherein the fluorescent protein is green fluorescent protein.

27. The method of claim 25, wherein the cell is a cell of the cell line designated 3134 deposited with American Type Culture Collection under accession number CRL-11998 (ATCC).

28. A method of monitoring the balance between levels of an agonist of a steroid receptor and an antagonist of the steoid receptor in a subject comprising:

a. periodically obtaining a biological sample from the subject,

b. contacting the sample with the cell of claim 5, and

c. directly detecting the location of fluorescence within the cell,

an increase in fluorescence aggregated at the site of the steroid receptor response element in the nucleus in a later-obtained sample relative to an earlier-obtained sample indicating an increase in level of the steroid agonist relative to level of the steroid antagonist in the sample, and a decrease in fluorescence aggregated at the site of the steroid receptor response element in the nucleus in a later-obtained sample relative to an earlier-obtained sample indicating an increase in level of the steroid antagonist of the steroid receptor relative to level of the steroid agonist in the sample.

29. A method of determining an effective dosage of a steroid receptor agonist in a subject comprising:

a. transferring into a set of cells from the patient a nucleic acid encoding a chimeric protein comprising a fluorescent protein fused to a steroid receptor;

b. contacting the cells in the set with one of a selected range of dosages of the steroid agonist; and

c. directly detecting location of fluorescence in the set of cells,

a dosage capable of locating fluorescence substantially in the nucleus indicating an effective dosage of steroid receptor agonist.

30. The method of claim 28, wherein the fluorescent protein is green fluorescent protein.

31. A method of determining an effective dosage of a steroid receptor agonist to maintain steroid receptor activation for a selected period of time in a subject comprising:

a. administering to the subject a dosage of the steroid receptor agonist,

b. periodically obtaining a biological sample from the subject,

c. contacting the sample with the cell of claim 5, and

d. directly detecting the location of fluorescence within the cell,

a dosage that maintains the location of fluorescence at the site of the steroid receptor response element array in the nucleus for the selected period of time indicating an effective dosage.

32. The method of claim 31, wherein the fluorescent protein is green fluorescent protein.

33. The method of claim 31, wherein the cell is a cell of the cell line designated 3134 deposited with American Type Culture Collection under accession number CRL-11998 (ATCC).

34. A method of determining an effective dosage of a steroid receptor antagonist to abrogate agonist activity for a selected period of time in a subject comprising:

a. administering to the subject a dosage of the steroid receptor agonist,

b. periodically obtaining a biological sample from the subject,

c. contacting the sample with the cell of claim 5, and

d. directly detecting the location of fluorescence within the cell,

e. a dosage that prevents the location of fluorescence at the site of the steroid receptor response element array in the nucleus for the selected period of time indicating an effective dosage.

35. The method of claim 34, wherein the fluorescent protein is green fluorescent protein.

36. The method of claim 34, wherein the cell is a cell of the cell line designated 3134 deposited with American Type Culture Collection under accession number CRL-11998 (ATCC).

37. A method of detecting a defect in a response pathway of a steroid receptor in a subject comprising transferring into a cell from the subject a nucleic acid functionally encoding a chimeric protein comprising a fluorescent protein fused to the steroid receptor and detecting the location of fluorescence within the cell as compared to the location of fluorescence within a normal, control cell transfected with the nucleic acid,

a difference in location of fluorescence within the cell of the subject as compared to location of fluorescence within the normal, control cell indicating a defect in the response pathway of the steroid receptor.

38. A method of determining whether a defect in a response pathway of a steroid receptor in a subject is in translocation of the steroid receptor to a cell nucleus, comprising transferring into a cell from the subject having the defect a nucleic acid functionally encoding a chimeric protein comprising a fluorescent protein fused to the steroid receptor and detecting the location of fluorescence within the cell,

the location of fluorescence substantially in the cytoplasm of the cell indicating the defect is in translocation of the steroid receptor to the nucleus.

39. A method of characterizing a ligand's effect on cellular localization of a compound to which the ligand binds in a cell comprising:

a. contacting the ligand with a cell having the compound fluorescently labeled by a fluorescent protein and

b. directly detecting the location of fluorescence within the cell,

the location of fluorescence in the cell indicating the localization effect of the ligand on the compound.

40. A method of determining a binding site for a DNA-binding protein comprising:

a. contacting the DNA-binding protein fluorescently labeled by a fluorescent protein with a cell having a plurality of copies of a nucleic acid having a putative binding site in an array such that the putative binding site can be directly visualized when bound by the fluorescently labeled DNA-binding protein, and

b. directly detecting the location of fluorescence within the cell, the presence of fluorescence aggregated at the putative binding site indicating a binding site to which the DNA-binding protein binds.

41. A chimeric protein comprising a fluorescent protein fused to a transcription factor.

42. The protein of claim 41, wherein the transcription factor is a steroid receptor.

43. The protein of claim 41, wherein the fluorescent protein is a green fluorescent protein.

44. The protein of claim 43, wherein the green fluorescent protein is *Aequorea* green fluorescent protein.

45. The protein of claim 43, wherein the green fluorescent protein has a substitution of threonine for serine at amino acid 65.

46. The protein of claim 43, wherein the green fluorescent protein is fused to the transcription factor via a peptide linker.

47. The protein of claim 46, wherein the peptide linker comprises about five glycine-alanine repeating units.

48. The protein of claim 42, wherein the steroid receptor is glucocorticoid receptor and a green fluorescent protein is fused to the amino-terminal end of the glucocorticoid receptor.

49. The protein of claim 48, wherein the glucocorticoid receptor has a substitution of serine for cysteine 656.

50. The protein of claim 49 having the amino acid sequence set forth in SEQ ID NO: 2.

51. An isolated nucleic acid encoding the protein of claim 41.

52. An isolated nucleic acid encoding the protein of claim 42.

53. An isolated nucleic acid encoding the protein of claim 43.

54. An isolated nucleic acid encoding the protein of claim 48.

55. An isolated nucleic acid encoding the protein of claim 50.

56. An isolated nucleic acid encoding the protein of claim 50 having the nucleotide sequence set forth in SEQ ID NO: 1.

57. A cell containing the nucleic acid of claim 51.

58. A cell containing the nucleic acid of claim 55.

ABSTRACT OF THE DISCLOSURE

The present invention provides a method of screening for a compound that binds to a selected nucleic acid comprising contacting compound fluorescently labeled by a fluorescent protein with a cell having a plurality of copies of the nucleic acid in an array such that the nucleic acid can be directly detected when bound by fluorescently labeled compound; and directly detecting the location of fluorescence within the cell,fluorescence aggregated at the site of the nucleic acid array indicating a compound that binds to the selected nucleic acid. In particular compounds such a transcription factors can be screened. Reagents for such method are provided including a mammalian cell having a plurality of steroid receptor response elements in an array such that the response element can be directly detected when bound by fluorescently labeled steroid receptor and a chimeric protein comprising a fluorescent protein fused to a steroid receptor.

FIG. 1

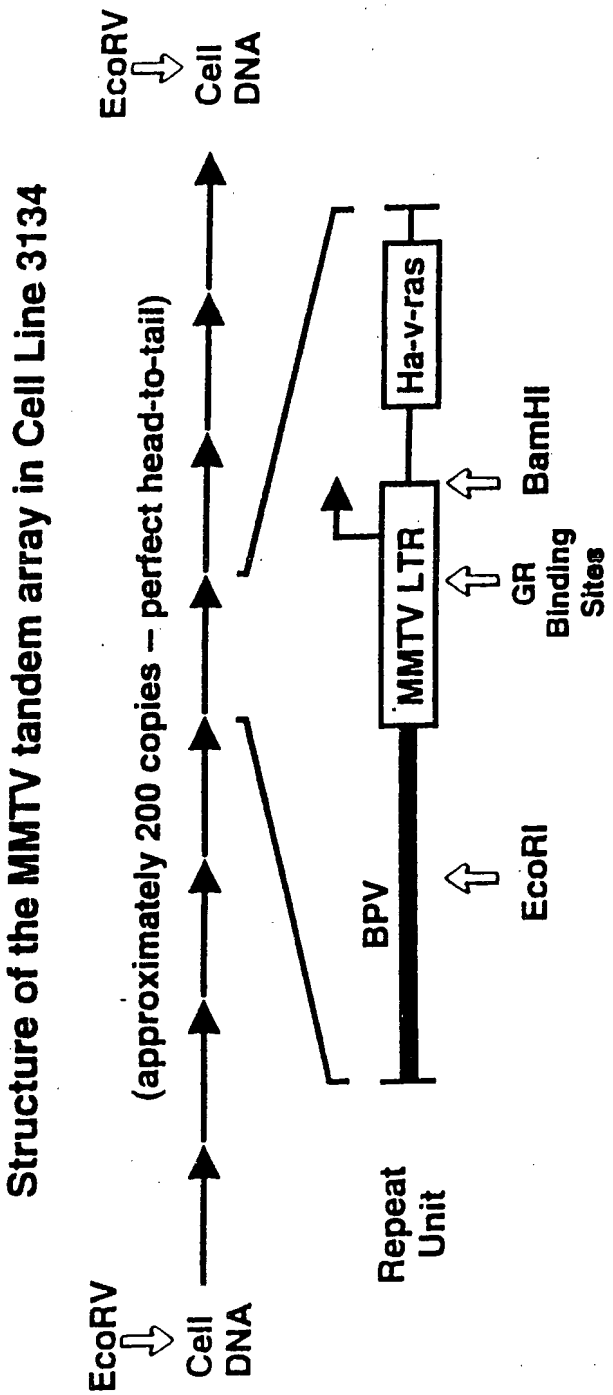

Structure of the MMTV tandem array in Cell Line 3134
EcoRV
Cell DNA
(approximately 200 copies – perfect head-to-tail)
EcoRV
Cell DNA
Repeat Unit
BPV
EcoRI
MMTV LTR
GR Binding Sites
BamHI
Ha-v-ras

FIG. 2

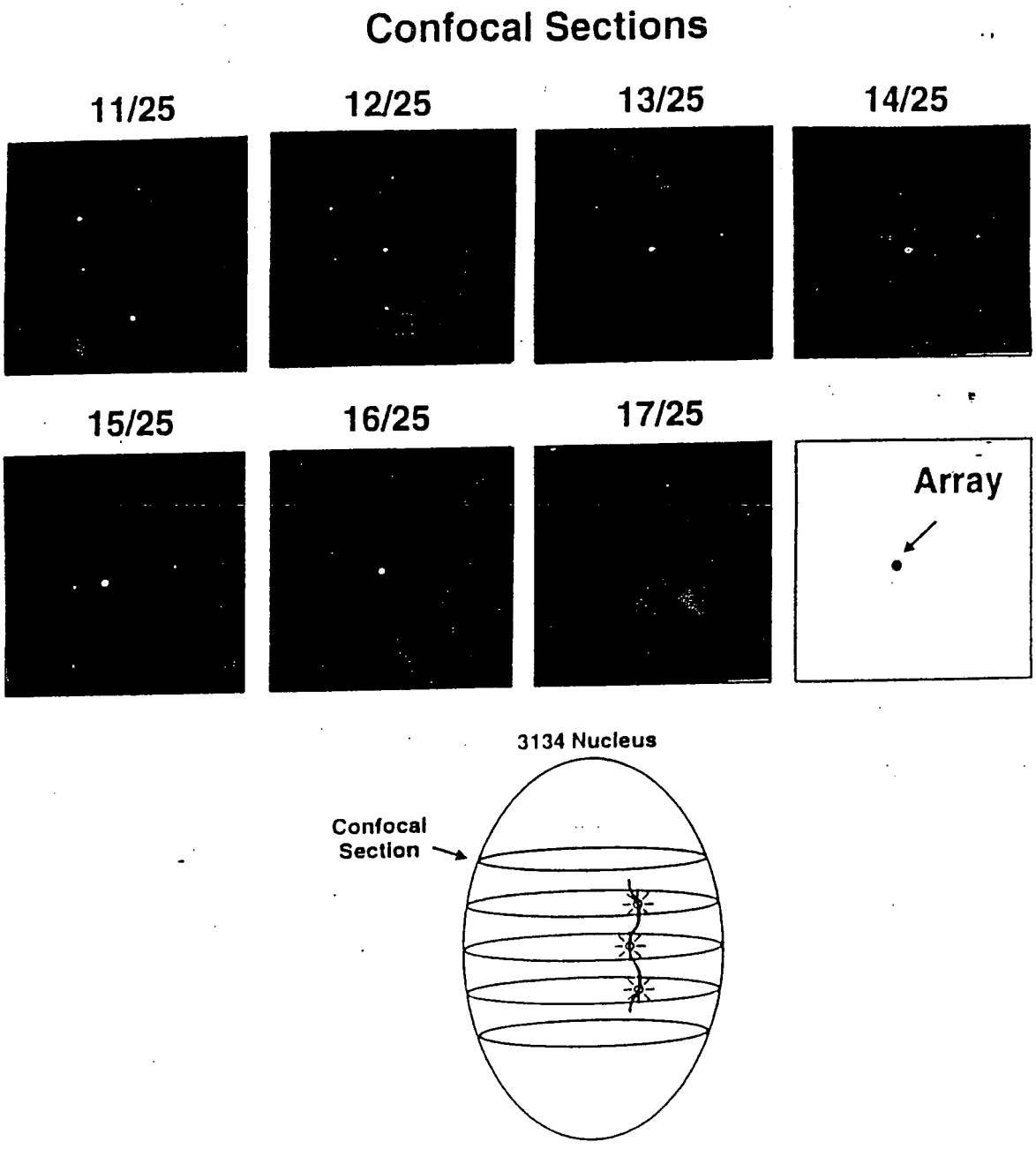
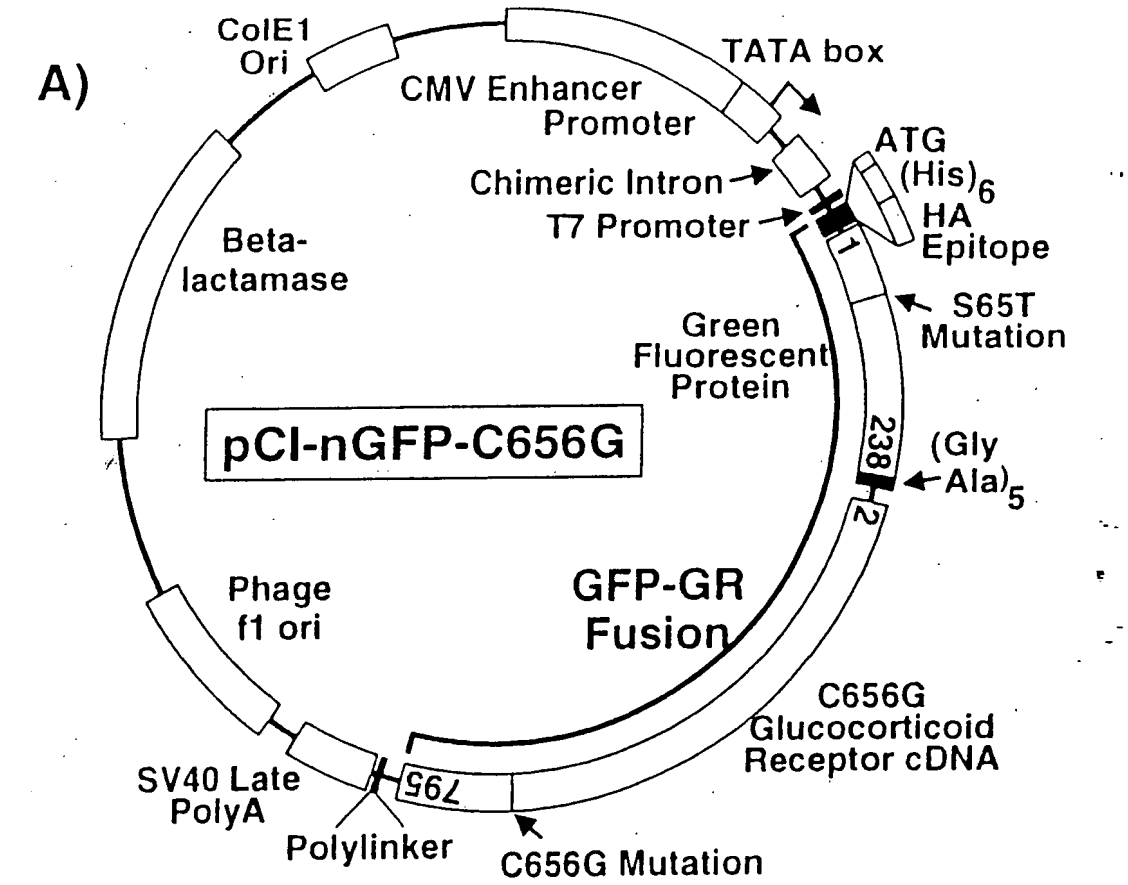
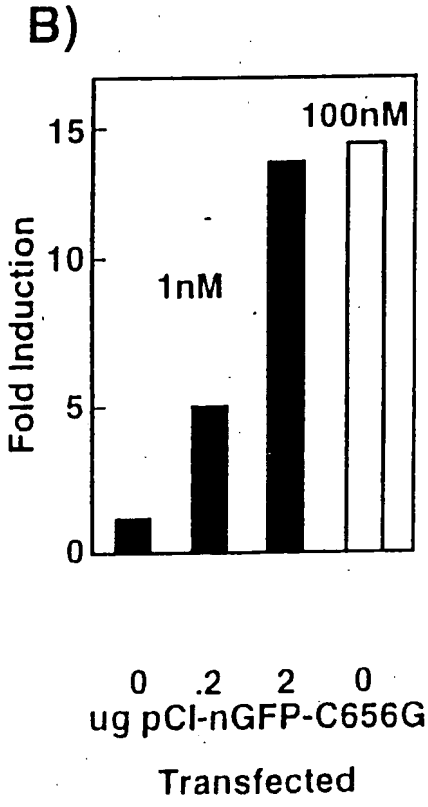

Confocal Sections
11/25
12/25
13/25
14/25
15/25
16/25
17/25
Array
3134 Nucleus
Confocal Section

FIG. 3


A)
ColE1 Ori
TATA box
CMV Enhancer Promoter
ATG
(His)6
HA Epitope
Chimeric Intron
T7 Promoter
Beta-lactamase
S65T Mutation
Green Fluorescent Protein
pCI-nGFP-C656G
(Gly Ala)5
Phage f1 ori
GFP-GR Fusion
C656G Glucocorticoid Receptor cDNA
SV40 Late PolyA
Polylinker
C656G Mutation
1
2
238
795


B)
Fold Induction
15
10
5
0
1nM
100nM
0 .2 2 0
ug pCI-nGFP-C656G
Transfected

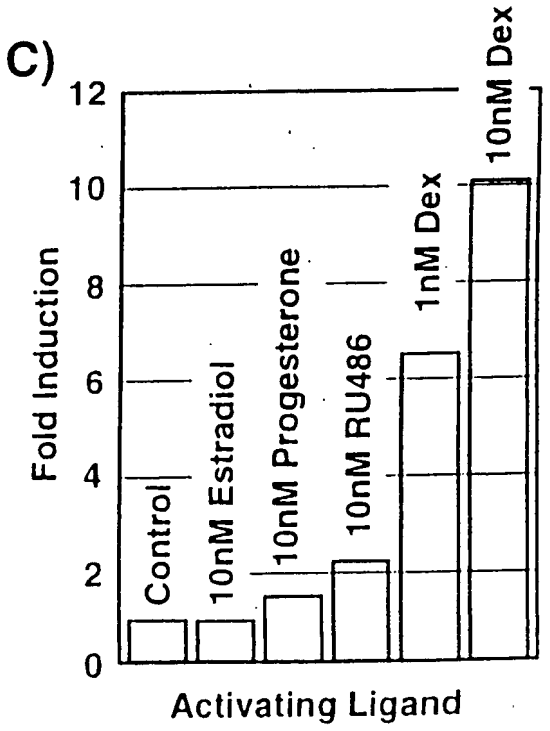
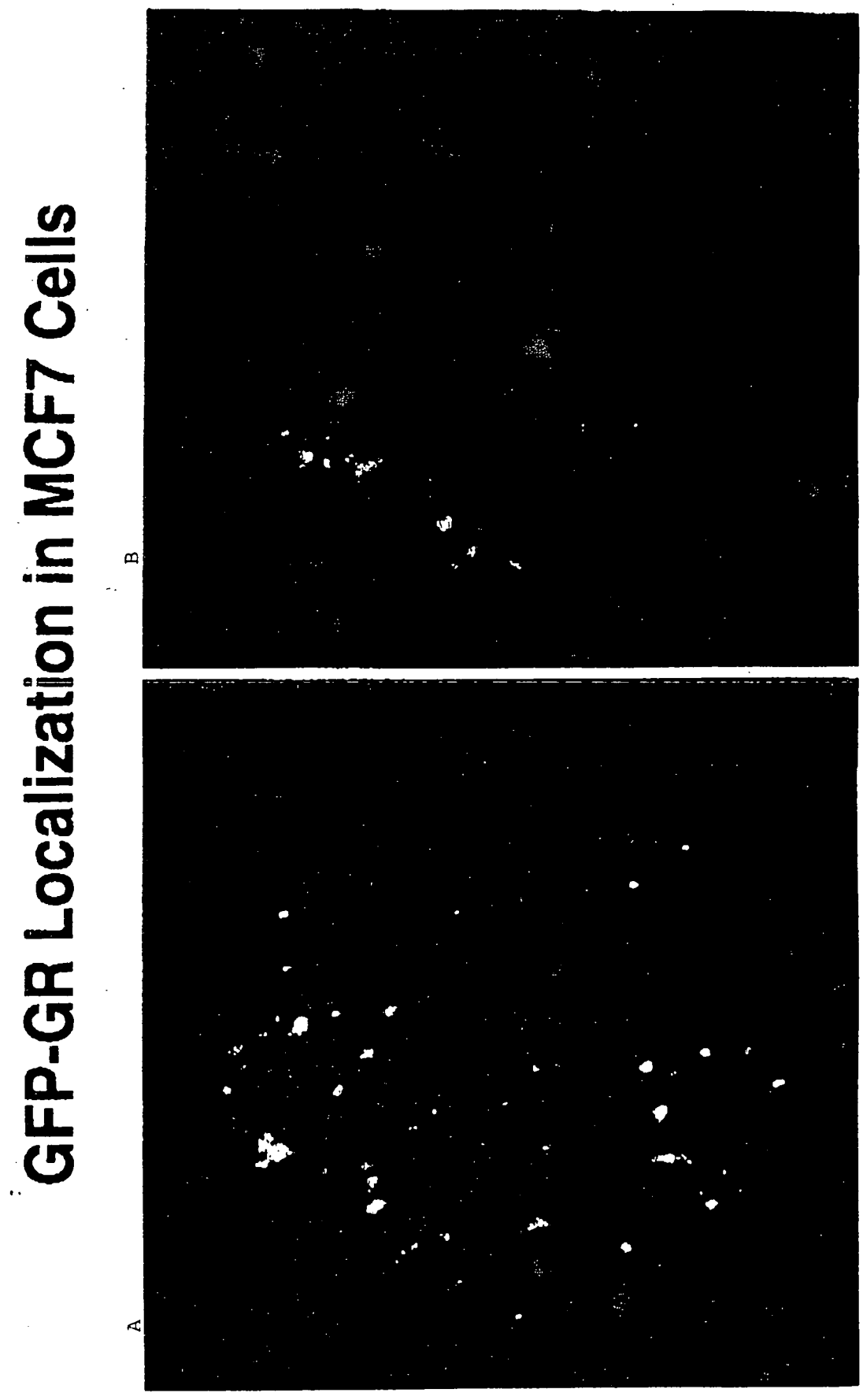
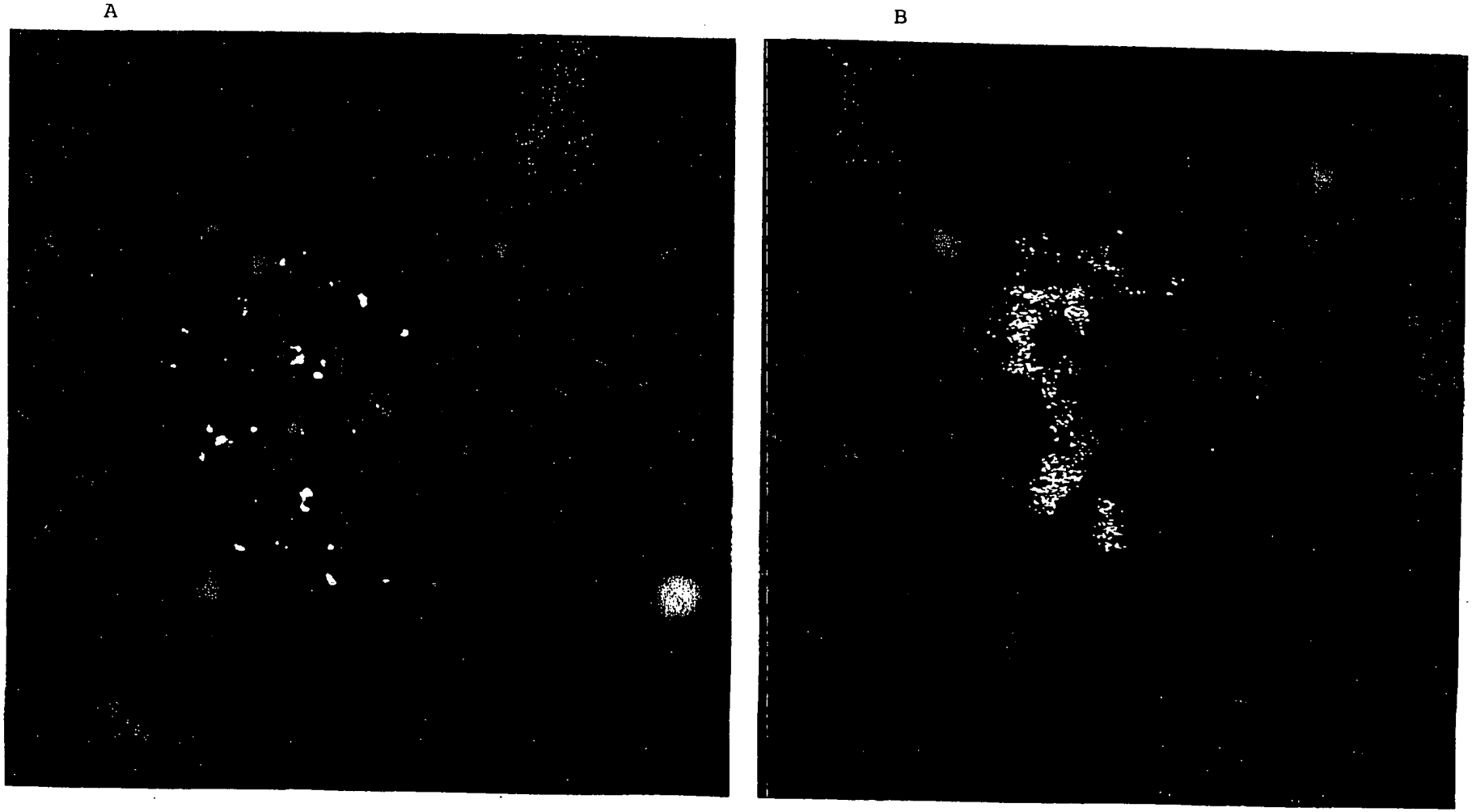

C)
Fold Induction
12
10
8
6
4
2
0
Control
10nM Estradiol
10nM Progesterone
10nM RU486
1nM Dex
10nM Dex
Activating Ligand

FIG. 4



Nuclear Localization of GFP-ER



MCF7

MDA-MB-231

5/5
FIG. 5

112

EVIDENCE APPENDIX 2

U.S. Serial No. 10/001,486
Monaco et al., Trends in Biotechnology, 1994, 12:280-286

YACs, BACs, PACs and MACs: artificial chromosomes as research tools

Anthony P. Monaco and Zoia Larin

Yeast artificial chromosomes (YACs) have become essential research tools as they enable large fragments of DNA to be cloned. In order to overcome several disadvantages of YACs, including chimaerism and instability, several complementary bacterial artificial chromosome (BAC) vectors have been developed. More recently, attempts are being made to construct artificial chromosomes in mammalian cells (MACs).

Artificial chromosomes have now become a major research tool in both genome analysis and in the functional characterization of genes. Yeast artificial chromosomes (YACs) (Ref. 1; see Glossary) have led the way in mapping complex genomes, and large YAC contigs (contiguous sets of overlapping clones) now cover much of the human genome[2]. YACs have been supplemented by other physical-mapping vectors, such as cosmids, bacteriophage P1 clones, bacterial artificial chromosomes (BACs) and, more recently, P1-derived artificial chromosomes (PACs) (see Glossary). For functional-gene analysis, YACs containing genes have complemented mutant phenotypes after transfer to mammalian cells and transgenic mice.

Mammalian artificial chromosomes (MACs) do not exist yet, but the elements needed to construct MACs are currently being defined (i.e. telomeres, centromeres and origins of replication) (see Glossary). MACs will be useful for understanding the structure of mammalian chromosomes and will provide vectors for the functional analysis of genes and gene complexes for gene transfer and somatic gene therapy. This review describes the development of artificial chromosomes, their use as research tools and the potential for constructing MACs.

Development of artificial chromosomes in yeast

Artificial chromosomes in the budding yeast *Saccharomyces cerevisiae* were developed following the isolation of the sequence elements for origins of replication[3], telomeres[4], and a centromere[5]. In 1983, these elements were combined together on the same linear vector and introduced into yeast cells, creating the first artificial chromosomes[6]. The chromosomal size requirements for normal replication and segregation in yeast were studied, and it was found that chromosome length was an important factor[6,7]. This pioneering work led Burke *et al.*[1] to generate a vector for cloning large human DNA molecules as linear artificial chromosomes in yeast (YACs). They showed that YACs could be used to generate whole libraries from the genomes of higher organisms with insert sizes at least ten-fold larger than those that could be accommodated by bacteriophage and cosmid vectors in bacterial hosts. Therefore, YACs bridged the gap that had existed between the insert capacity of previous cloning vectors and the ability to construct physical maps of higher organisms.

YACs in physical-mapping and positional-cloning projects

YACs quickly became the major cloning vector in use in a variety of genome projects ranging from bacteria and plants to *Drosophila, Caenorhabditis elegans*, mouse and human[8]. Their use became widespread, not only because they could accommodate large inserts of DNA, so that fewer clones were required to cover a particular genome, but also because YAC libraries gave better coverage of the genome than previously had been observed with cosmid vectors. The improved genome representation has substantially aided the *C. elegans* genome project in that YACs enabled previously unlinked cosmid contigs to be physically joined[9]. This provided long-range continuity for the *C. elegans* genome, enabled the physical and genetic maps to be aligned, and a minimum set of overlapping YAC clones on replica filters to be distributed.

A. P. Monaco is at the Imperial Cancer Research Fund Laboratories, Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford, UK OX3 9DU. Z. Larin is at the CRC Chromosome Molecular Biology Group, Department of Biochemistry, University of Oxford, South Parks Road, Oxford, UK OX1 3QU.

Glossary

Alphoid satellite DNA – A tandemly repeated DNA sequence present at human centromeres, comprising a basic monomeric repeat of 170 bp. This small repeat is organized into higher-order units that have been shown to be specific to one, or a small group of, human chromosomes.

BAC – Bacterial artificial chromosome; a cloning sytem for isolating genomic DNA based on the F-factor plasmid, the bacterial sex or fertility plasmid, which has a low copy number because of the strict control of replication.

Centromere – The region of the chromosome that is constricted and is the site of attachment of the spindle during meiosis or mitosis. It is necessary for the stability and proper segregation of chromosomes during meiosis and mitosis, and was first isolated in yeast. It is an essential component of yeast artificial chromosomes.

Chimaerism – The term used to describe a DNA insert in a cloning vector that is from two independent sources or chromosome regions and, thus, does not represent a contiguous stretch of DNA. An example of a chimaeric YAC clone is shown in Fig. 1.

Chromosome walking – The isolation of sequential pieces of genomic DNA in a set of clones to form a contiguous overlapping set (a contig). This is usually accomplished by isolating the end-fragment of one clone, which is used to rescreen the library for the next overlapping clone. This process (walking) is repeated along the chromosome until the gene of interest is reached.

Cosmid – A plasmid cloning vector used for isolating genomic DNA that contains a bacteriophage lambda cos site that directs insertion of DNA into phage particles.

Euchromatin – Those regions of the genome that are not always condensed and contain expressed genes.

FISH – Fluorescence *in situ* hybridization; a technique that visualizes the hybridization of a nucleic acid probe with a fluorescent tag to RNA or DNA in cells, tissues or chromosomes on slides. In physical mapping, it is used to localize genes or genomic clones to particular cytogenetic bands of a chromosome, or to infer order of clones along the chromosome.

Fragile site – A chromosome locus that separates, producing a gap that is visible cytogenetically. Fragile sites are usually inducible after treatment of cultured cells in media with certain components, such as pyrimidine triphosphate reducing agents, as seen in the Fragile X mental retardation site in Xq27.3.

Heterochromatin – Those regions of the genome that are condensed and are not considered to have expressed genes. On most chromosomes, the centromere is considered to be constitutive heterochromatin and comprises tandemly repeated DNA sequences such as alphoid satellite DNA. The inactive X chromosome is considered to be facultative heterochromatin as it is not always in this condensed state.

LINE – Long interspersed nucleotide elements that are retroposons of RNA polymerase II transcripts in mammalian genomes. In humans, the most common LINE repeat has a full-length sequence of almost 6 kb and is thought to encode a reverse transcriptase.

Lipofection – A method of DNA transfer to cells in tissue culture or *in vivo* where the gene of interest is mixed with a cationic lipid suspension and applied to the cells of interest.

Metaphase spread – A preparation of chromosomes, on a microscope slide, in their condensed phase, usually seen at metaphase. Cells are cultured for synchronization of the mitotic cycle to achieve long chromosomes and the highest percentage of cells at metaphase at the time of harvesting.

Origin of replication – Site of initiation of DNA synthesis; first described in yeast as the autonomous replication sequence (ARS) and is an important element for the replication of yeast artificial chromosomes.

PAC – P1-derived artificial chromosome; a cloning system for isolating genomic DNA based on the F-factor plasmid, as in BACs, but also containing some of the elements of the bacteriophage P1 cloning system.

P1 clones – A cloning system for isolating genomic DNA that uses elements from bacteriophage P1 (i.e. recombination sites, loxP and a packaging site, pac).

Replicon – The segment of the genome in which DNA is replicated and, by definition, contains an origin of replication.

Telomere – The end of a chromosome comprising simple repeat DNA that is synthesized by a ribonucleoprotein enzyme called telomerase. Telomeres from *Tetrahymena* were the first to be isolated and shown to be functional in yeast; thus, they are used to seed telomere formation at the ends of yeast artificial chromosomes.

SINE – Short interspersed nucleotide element that is highly repeated in the genome. It comprises retroposons from RNA polymerase III transcripts. In humans, the Alu sequence is the most common SINE repeat and is ~300 bp long.

YAC – Yeast artificial chromosome; a cloning system in yeast comprising an autonomously replicating linear vector containing an exogenous DNA insert, flanked by a yeast centromere and two telomere seeding sequences.

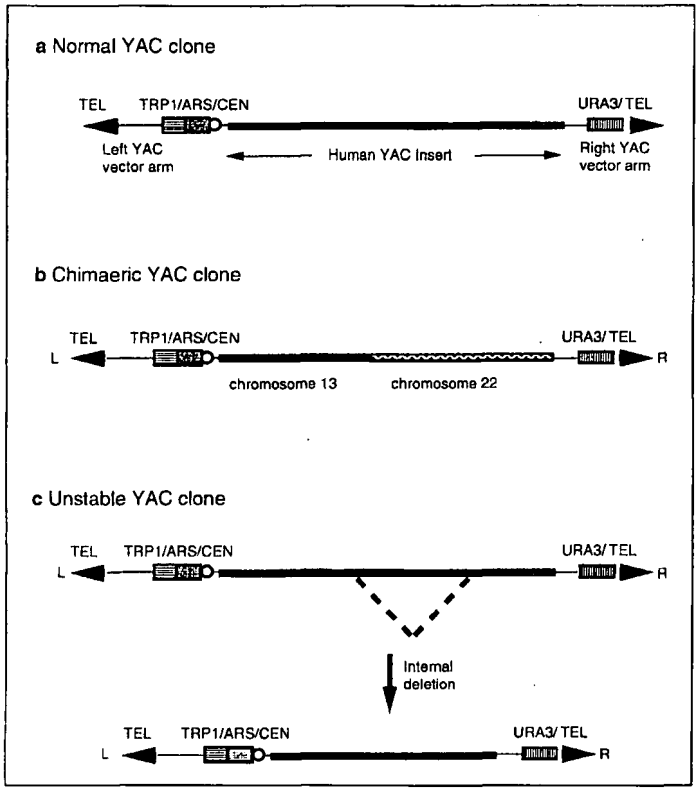


Figure 1

(a) Normal yeast artificial chromosome (YAC) clone with *Tetrahymena* telomeres (TEL) at either end, yeast selectable markers TRP1 and URA3, a yeast origin of replication (ARS, autonomously replicating sequence), a yeast centromere (CEN), and a large human DNA insert. (b) A chimaeric YAC clone with the human DNA insert comprising unlinked DNA segments; for example, from chromosomes 13 and 22, together in the same clone. (c) An unstable YAC clone that undergoes rearrangement and deletion of internal sequences.

In the human genome project, YACs have been used to map large chromosome segments. There has been early success in targeting Xq24–qter, a region of the X chromosome that is interesting because it contains two fragile sites (FRAXA and FRAXE) (see Glossary) and has a high density of genes in Xq28 (Ref. 8). Using a combination of whole-genome YAC libraries and human YACs derived from a somatic-cell hybrid containing this region, almost complete coverage of Xq24–qter has been achieved by hybridization screening with probes and chromosome walking (see Glossary). More recently, two human chromosomes (21q and euchromatic Y) (see Glossary) have been covered by YAC contigs using total genomic YAC libraries and polymerase chain reaction (PCR)-based screening with chromosome-specific sequence-tagged sites (STSs) (Refs 10–12).

Cohen *et al.*[2] have described a first-generation physical map of the entire human genome using a combination of STS screening of very-large-insert YAC libraries with 2000 Généthon genetic markers[13] and gel-fingerprinting of 33 000 YAC clones by hybridization with human repeat sequences[14]. The Généthon genetic markers are highly polymorphic microsatellites of tandemly repeated dinucleotides flanked by unique primers that enable genomic DNA samples to be tested by PCR. As these markers have been used to generate a genetic map of the whole human genome[13], screening the Centre d'Etude du Polymorphisme Humain (CEPH) YAC libraries with them has enabled the genetic and physical maps to be merged.

The rapid progress made in obtaining a physical map of the human genome is likely to have its largest impact on the ability to isolate genes responsible for inherited disorders when there is no prior knowledge of the defective protein. This process is termed 'positional cloning', because the position of the defective gene in the genome is essential to its eventual isolation and characterization[15]. YACs have proved to be an important tool for isolating large regions of DNA around the linkage-map positions of inherited diseases. This has been most fruitful when an abnormality of chromosome structure is associated with some cases of the disease (for example, fragile sites, trans-locations or deletions).

Fluorescence *in situ* hybridization (FISH) (see Glossary) of YAC clones on metaphase spreads (see Glossary) of patient chromosomes has enabled the precise locations of several genes that were structurally disrupted, giving rise to the disease, to be found[16,17]. When YAC clones have been isolated close to the position of the defective gene on the chromosome, candidate genes are then isolated using a variety of positional-cloning techniques[15].

Alternative cloning vectors: P1, BACs and PACs

There are several major problems that need to be considered when working with YAC clones (Fig. 1). One is the high percentage (40–60%) of clones (particularly in human YAC libraries) that are chimaeric (see Glossary), i.e. they contain two separate segments of unlinked human DNA in the same YAC clone. This may be due to co-ligation of DNA inserts *in vitro* prior to yeast transformation, or to recombination *in vivo* between two DNA molecules that were introduced into the same yeast cell. Chimaerism causes problems in chromosome-walking and gene-isolation experiments but, usually, can be detected by methods such as FISH analysis on metaphase chromosomes or YAC end isolation and subsequent mapping. Isolating the ends of YAC clones by PCR-based or subcloning methods is important to extend YAC contigs by rescreening the YAC libraries for adjacent YACs along the chromosome (i.e. chromosome walking). A second problem with YACs is that some clones are unstable and tend to delete internal regions from their inserts. This can be deleterious when constructing physical maps of chromosome regions or in gene-isolation experiments. The instability and chimaerism of YACs can be decreased, to some extent, by transferring YAC clones to, or constructing YAC libraries in, recombination-deficient strains[18,19].

The third major problem with YAC clones is that the 15 Mb yeast host chromosome background cannot be separated from the YACs by simple methods.

In contrast with bacterial-based cloning vectors, where it is relatively easy to purify the plasmid vector and insert DNA from the bacterial host DNA, YACs have a very similar structure to the natural yeast chromosomes. Purifying YAC inserts from the natural yeast chromosomes usually requires separation by pulsed-field gel electrophoresis (PFGE) or direct subcloning of the entire yeast genome into bacteriophage or cosmid vectors, followed by identification of human-specific clones derived from the YAC.

In the past few years, several alternative cloning systems based on bacterial host systems have been developed. These systems may have several advantages over YACs (Table 1), including a lower frequency of chimaerism, a higher transformation efficiency in generating libraries, and ease of purifying insert DNA away from the host genome. However, the size of insert that can be cloned with these alternative systems is smaller compared with YACs (Table 1), and they have not, as yet, been characterized as thoroughly as YACs for problems such as instability.

Sternberg[20] has developed a bacteriophage P1 cloning system that is able to clone inserts of up to 100 kb in size. This system packages both cloned DNA and vector into phage particles, injects them into *Escherichia coli* and circularizes the DNA using the P1 loxP recombination sites and a host expressing the P1 Cre recombinase. The vector carries a gene for kanamycin resistance for selection and a P1 plasmid replicon (see Glossary) for maintaining the vector at one copy per cell (thus avoiding high-copy-number instability of clones), but enabling the DNA to be amplified before re-isolation. Both human and mouse genomic P1 libraries have been constructed for screening[21,22].

Another cloning system in *E. coli* is the bacterial artificial chromosome (BAC) (Ref. 23). This system uses an F-factor-based vector and can accommodate inserts up to 300 kb. As the vector is a low-copy-number plasmid, the clones appear to be stable over many generations; however, this has not been tested for different regions of the genome. BACs can be transformed into the *E. coli* host very efficiently by electroporation, thus avoiding the packaging extracts that are required with the P1 system. However, the disadvantages of the BAC system include the lack of positive selection for clones containing inserts and a very low yield of DNA.

More recently, Ioannou *et al.*[24] have developed the P1-derived artificial chromosome (PAC) cloning system. They constructed a new vector that incorporates features of both the P1 and F-factor systems and can be transformed into the *E. coli* host by electroporation. The PAC vector is able to handle inserts in the 100–300 kb range and, as yet, there are no major problems with chimaerism or clone instability in the one initial report describing PACs. These new vectors, along with the more traditional cosmid vectors, may be important in complementing YAC contigs for gene isolation and sequencing projects in the future.

Table 1. Comparison of cloning vectors

Vector	Host	Structure	Insert size
Cosmids	*E. coli*	Circular plasmid	35–45 kb
P1 clones	*E. coli*	Circular plasmid	70–100 kb
BACs	*E. coli*	Circular plasmid	up to 300 kb
PACs	*E. coli*	Circular plasmid	100–300 kb
YACs	*S. cerevisiae*	Linear chromosome	100–2000 kb
MACs	Mammalian cells	Linear chromosome	? >1000 kb

Functional studies using YACs

One of the major advantages of YACs is their ability to contain the complete genomic locus of large genes, including a range of alternative promoters and exons and their upstream control elements. This has made YACs attractive vectors for functional gene studies after transfer from yeast to mammalian cells[25]. Recently, YACs have been used by several groups to generate transgenic mice to complement mouse mutations[26]. It is also feasible that YAC transgenes could be used in the positional cloning of genes for mouse mutations, based on the ability of the input DNA to complement the phenotype. Before transferring a YAC to mammalian cells, a suitable selectable marker, such as the gene for neomycin resistance, must be introduced onto the YAC (Fig. 2). This has been done by researchers in several laboratories using recombination in yeast after transformation with both integration and with replacement vectors that have homology to the YAC vector or to internal repeat sequences such as SINEs (Alu) or LINEs (see Glossary)[28–33]. Recombination in yeast can also be used to delete sequences from either end of a YAC using 'fragmentation' vectors[33] (Fig. 2), or to reconstruct two smaller overlapping YACs into a larger YAC (Ref. 12).

Once a YAC of interest has been 'retrofitted' with an appropriate mammalian selectable marker, there are several choices for transferring the vector to mammalian cells in culture (for review see Ref. 25). One

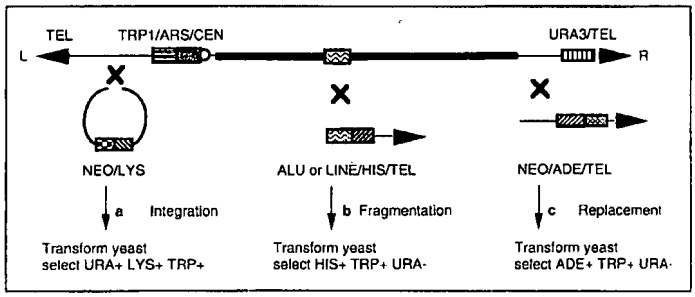


Figure 2

Three different ways of 'retrofitting' a YAC clone with (**a**) an integration vector to put the neomycin resistance gene (NEO) and a yeast selectable marker (LYS) into the YAC left arm (L); (**b**) a fragmentation vector that truncates the YAC after recombination with human repeat sequences such as ALU or LINE, using a vector with a cloned *Tetrahymena* telomere; and (**c**) a vector to put the NEO gene on the right arm of the YAC after replacing the normal URA selectable marker. (Reprinted, with permission, from Ref. 27.)

117

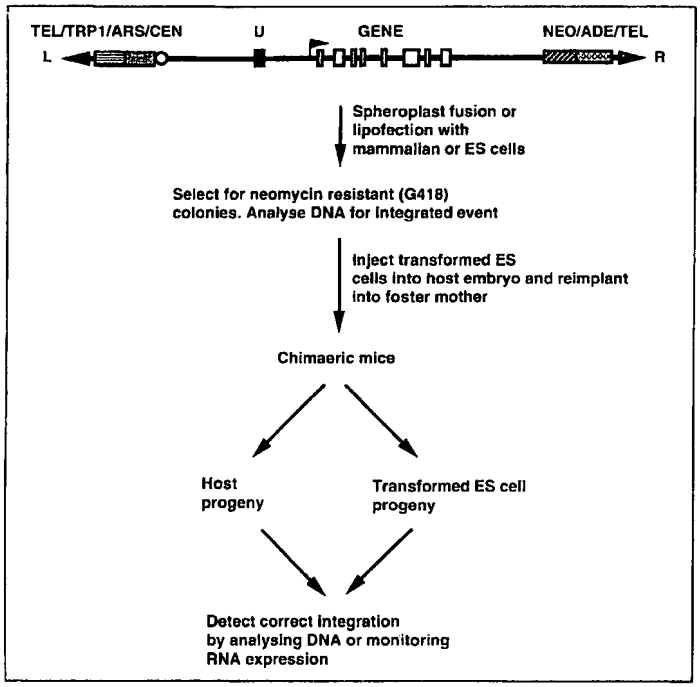


Figure 3

A normal YAC clone retrofitted with the neomycin resistance gene (NEO) and containing the complete genomic locus for a large gene with its upstream control elements (U). This YAC can be transferred to mammalian cells, including embryonic stem (ES) cells, by methods such as lipofection or spheroplast fusion. Cells that have the YAC integrated are resistant to G418 and can be analysed for integrity of the YAC and the gene of interest. For generating transgenic mice, transformed ES cells are injected into blastocysts to produce chimaeric mice which, in turn, are bred to produce homozygous YAC transgenic mice. Further analysis of these animals will indicate whether the transgene is appropriately expressed and if the mutant phenotypes have been complemented. (Reprinted, with permission, from Ref. 27.)

method is to remove the yeast cell wall and to fuse the yeast spheroplasts with mammalian cells using polyethylene glycol. There is no physical isolation of DNA during the fusion, so large YACs can be transferred intact. One drawback is the simultaneous transfer of the complete yeast genome, which usually integrates with the YAC into the same site in the mammalian host genome. This method has been very successful in transferring YACs intact to cultured hamster and mouse cells, but not to human cells. Another approach is to physically isolate the YAC insert using PFGE. The YAC is excised from the gel, the agarose digested and the YAC, in solution, transferred to the mammalian cells by microinjection or lipofection-based procedures (see Glossary). It is important when purifying large YAC inserts to use procedures that avoid shearing the DNA.

The generation of transgenic mouse lines with YACs has been successful using both pronuclear injection of purified YAC DNA and by introduction into embryonic stem (ES) cells followed by generation of chimaeric mice (for review see Ref. 26). In the latter approach, introduction of YACs into ES cells has been accomplished by yeast spheroplast fusion, or by lipofection of purified YAC DNA (Fig. 3). The advantage of the ES-cell approach is that the integrity of the transferred YAC can be studied in the ES cells prior to injection into blastocysts for the production of chimaeric mice. However, it is more laborious than direct pronuclear injection as it requires germline transmission of the YAC transgene. The production of YAC transgenic mice should prove to be very important for studying large and complex genes or gene clusters and, combined with the ability to alter the gene by recombination in yeast prior to transgenesis, should help to dissect gene regulation and function.

Mammalian artificial chromosomes

The success in the construction of artificial chromosomes in yeast has spurred interest in isolating the elements involved in the structure of mammalian chromosomes. If all the elements were identified, then artificial chromosomes in mammalian cells could be constructed (for review see Ref. 34). Mammalian artificial chromsomes (MACs) would provide an experimental system for defining the size requirements for accurate mitotic and meiotic segregation, and for studying chromosome function in mammalian cells. In addition, they would provide an autonomously replicating transfer vector for functional analysis of large complex genes and a potential tool for somatic gene therapy.

Construction of artificial chromosomes in yeast was dependent on the ability to isolate, and assay for, elements such as origins of replication, telomeres and a centromere. Telomeres have been successfully isolated from human DNA by complementation using modified YAC vectors containing just a single *Tetrahymena* telomere cassette[35]. However, human centromeres and origins of replication have proved more difficult to isolate. Human centromeres are much larger and more complex than those in yeast. The Y-chromosome centromere is the best-studied human centromere. It contains a large block of tandemly repeated alphoid satellite DNA (see Glossary) (240–1600 kb), flanked by several other minor satellite arrays and localized repetitive sequences[36,37].

Mammalian origins of replication have been much more difficult to define by genetic assays and have largely been studied by mapping the sites of initiation of DNA replication using two-dimensional gel systems. The data suggest that origins of replication in mammalian DNA may not be precise and may begin at multiple sites along the DNA, with some sites being stronger than others (for reviews see Refs 38–40). However, recent evidence from studying ~200 kb of the human β-globin-gene domain using a replication-direction assay indicates a discrete origin of replication[41]. This assay used the incorporation of bromodeoxyuridine (BrdU) into DNA to separate the newly synthesized leading-strand of DNA, followed by hybridization with strand-specific DNA probes. The direction of replication is determined by an unequal hybridization pattern when comparing probes from each strand.

MAC Construction

Two basic strategies for generating MACs are considered (for review see Ref. 42). One approach is to fragment mammalian chromosomes near the centromere using cloned human telomere sequences. The other approach is to define the minimum sequence requirements for each element, and to construct a MAC in yeast by joining these cloned elements.

Telomere-fragmentation experiments involve transfection of cloned human telomeres into mammalian cells, producing centric (containing a functional centromere) and acentric (lacking a functional centromere) fragments after chromosome breakage, and formation of new functional telomeres[42,43]. The efficiency of breakage, however, is poor, as deletion of the distal ends of chromosomes occurs preferentially. Chromosome breakage is more frequent when positive and negative counter-selection mechanisms are also in operation. For example, one can select positively for retention of a marker on the short arm of a chromosome and negatively for loss of the distal portion of the long arm of the same chromosome, thus enriching for telomere fragmentation between the positive and negative markers. Using this approach, Farr *et al.*[44] produced a series of defined terminal deletions of the long arm of the X chromosome. This work indicated that telomere fragmentation could be used as a complementary mapping strategy for other, less-well-defined chromosomes where there are few existing somatic-cell-hybrid mapping reagents.

An alternative strategy is to target defined sequences in association with human telomeres[45]. In this method, the frequency of breakage is dependent on the length of sequence homology during homologous recombination, and is significantly elevated when using isogenic (i.e. genetically uniform) DNA from the appropriate cells[46]. From these studies, it appears that telomere-directed fragmentation could be used to construct functional mini-chromosomes containing the minimum centromeric sequences and replication origins required for autonomous segregation in mammalian cells[42].

An alternative approach for constructing an artificial chromosome would be to assemble the essential components *in vitro*, or in yeast, and re-introduce them into mammalian cells. This strategy requires a much more detailed knowledge and understanding of the sequence elements required. The structure and function of mammalian telomeres is understood; however, progress has been slower in identifying mammalian centromeres and replication origins, for several reasons. First, the centromere of human chromosomes is a much larger region than the telomere. There are only a few kilobases of telomeric DNA but, possibly, a few hundred kilobases of centromeric DNA. The size of the region therefore necessitated isolating cloned centromeric DNA in YACs. However, these tandemly repeated sequences (alpha satellite or alphoid DNA) are unstable in yeast[18]. Second, defining the minimum size of the region and identifying the essential sequences by physical mapping procedures requires functional analysis *in vivo*.

Recent structural and functional studies of the human Y-chromosome centromere has furthered our understanding of these sequences. A Y-chromosome centromeric interval has been defined by analysis of rearranged Y chromosomes from different cell lines: ~500 kb containing a few hundred kilobases of alphoid DNA and short-arm flanking sequences is required for mitotic segregation[47]. A functional analysis of exogenous Y alphoid DNA after transfer to human and hamster cells has demonstrated that a few hundred kilobases is sufficient to direct *de novo* formation of several features of a centromere and initiate dicentric chromosome behaviour in human cells[48]. Further functional studies, including alphoid DNA with or without any short arm sequences, in association with human telomeres, would be required to determine whether or not these sequences are sufficient for autonomous mitotic centromere function. Finally, artificial chromosomes will require the ability to replicate but, unlike yeast origins of replication, our understanding of mammalian origins is limited[38–40], although replication origins probably occur relatively frequently within DNA (every few tens of kilobases). As size may also be an important factor for the stability of chromosomes, any large piece of appropriate DNA as a stuffer fragment may contain sufficient replication origins.

Although a combination of these elements may be assembled in yeast, difficulties may arise when transferring large molecules as MACs to mammalian cells. Methods to transfer YACs to mammalian cells include lipofection and yeast spheroplast fusion, but there are advantages and disadvantages with both. While lipofection may transfer purified YAC DNA molecules efficiently to most cells, large DNA molecules (>600 kb) may not be transferred intact; alternatively, fusion of yeast and mammalian cells does not limit the size of YAC DNA molecules transferred, although this has been successful in some cell lines and not others. Thus, these methods may not be suitable for transferring MACs, which may be much larger molecules. Other methods currently being used for somatic gene therapy for efficient delivery of genes, such as viral-mediated-adenovirus–polylysine–DNA complexes[49], may be more suitable, as they might not have problems with large DNA molecules, although this has not been tested.

MACs in somatic gene therapy and the generation of transgenic mice

The manipulation of MACs for somatic gene therapy offers an exciting alternative to existing strategies currently being tested for the treatment of human genetic diseases (for review see Ref. 50). There are, potentially, several advantages of using MACs compared with other methods. First, as MACs will replicate autonomously within the host cell, they will not function as insertional mutagens by inserting into the patient's genome. Second, because the transfer vector

will be large (i.e. possibly a few megabases), it will have the capacity to accommodate the entire repertoire of a large gene and its regulatory elements, which may itself span megabases of DNA. Third, some genetic diseases are a result of defects in more than one gene, termed contiguous gene-deletion syndromes, and have complex phenotypes, which are usually a combination of individual phenotypes for each gene that is deleted. As MACs will have the capability of accommodating an entire cluster of different genes, it may be possible to correct the respective phenotype by complementing several mutant loci at the same time, rather than individually.

Finally, the study of MAC function and stability during meiosis will be important. This will be achieved by generating transgenic mice from the transfection of ES cells with a MAC to produce chimaeric mice. Experimental approaches to analyse MACs in transgenic mice after germline transmission may provide insight into mammalian chromosome function.

Acknowledgements

We thank Chris Tyler-Smith and Ulrich Müller for comments on the manuscript.

References

1 Burke, D. T., Carle, G. F. and Olson, M. V. (1987) *Science* 236, 806–812
2 Cohen, D., Chumakov, I. and Weissenbach, J. (1993) *Nature* 366, 698–701
3 Struhl, K., Stinchomb, D. T., Scherer, S. and Davis, R. W. (1979) *Proc. Natl Acad. Sci. USA* 76, 1035–1039
4 Szostak, J. W. and Blackburn, E. H. (1982) *Cell* 29, 245–255
5 Clarke, L. and Carbon, J. (1980) *Nature* 287, 504–509
6 Murray, A. W. and Szostak, J. W. (1983) *Nature* 305, 189–193
7 Murray, A. W., Schultes, N. P. and Szostak, J. W. (1986) *Cell* 45, 529–536
8 Schlessinger, D. (1990) *Trends Genet.* 6, 248–258
9 Coulson, A., Kozono, Y., Lutterbach, B., Shownkeen, R., Sulston, J. and Waterston, R. (1991) *BioEssays* 13, 413–417
10 Chumakov, I. *et al.* (1992) *Nature* 359, 380–387
11 Foote, S., Vollrath, D., Hilton, A. and Page, D. C. (1992) *Science* 258, 60–66
12 Green, E. D. and Olson, M. V. *Science* 250, 94–98
13 Weissenbach, J. *et al.* (1993) *Nature* 359, 794–801
14 Bellanné-Chantelot, C. *et al.* (1992) *Cell* 70, 1059–1068
15 Collins, F. S. (1992) *Nature Genet.* 1, 3–6
16 Heitz, D. *et al.* (1991) *Science* 251, 1236–1239
17 Tümer, Z. *et al.* (1992) *Hum. Mol. Genet.* 1, 483–489
18 Neil, D. L., Villasante, A., Fisher, R. B., Vetrie, D., Cox, B. and Tyler-Smith, C. (1990) *Nucleic Acids Res.* 18, 1421–1428
19 Ling, L. L., Ma, N. S-F., Smith, D. R., Miller, D. D. and Moir, D. T. (1993) *Nucleic Acids Res.* 21, 6045–6046
20 Sternberg, N. (1990) *Proc. Natl Acad. Sci. USA* 87, 103–107
21 Sternberg, N., Ruether, J. and deReil, K. (1990) *New Biol.* 2, 151–162
22 Pierce, J. C., Strenberg, N. and Sauer, B. A. (1992) *Mamm. Genome* 3, 550–558
23 Shizuya, H. *et al.* (1992) *Proc. Natl Acad. Sci. USA* 89, 8794–8797
24 Ioannou, P. A. *et al.* (1994) *Nature Genet.* 6, 84–89
25 Huxley, C. and Gnirke, A. (1991) *BioEssays* 13, 545–550
26 Forget, B. G. (1993) *Proc. Natl Acad. Sci. USA* 90, 7909–7911
27 Larin, Z. (1993) *AgBiotech News Inf.* 5, 419N–423N
28 Pavan, W. J., Hieter, P. and Reeves, R. H. (1990) *Mol. Cell. Biol.* 10, 4163–4169
29 Srivastava, A. K. and Schlessinger, D. (1991) *Gene* 103, 53–59
30 Davies, N. P., Rosewell, I. R. and Brüggemann, M. (1992) *Nucleic Acids Res.* 20, 2693–2698
31 Riley, J. H., Morten, J. E. N. and Anand, R. (1992) *Nucleic Acids Res.* 20, 2971–2976
32 Markie, D. *et al.* (1993) *Somat. Cell Mol. Genet.* 19, 161–169
33 Pavan, W. J., Hieter, P. and Reeves, R. H. (1990) *Proc. Natl Acad. Sci. USA* 87, 1300–1304
34 Tyler-Smith, C. and Willard, H. F. (1993) *Curr. Opin. Genes Dev.* 3, 390–397
35 Brown, W. R. A. (1989) *Nature* 338, 774–776
36 Cooper, K. F., Fisher, R. B. and Tyler-Smith, C. (1992) *J. Mol. Biol.* 228, 421–432
37 Cooper, K. F., Fisher, R. B. and Tyler-Smith, C. (1993) *J. Mol. Biol.* 230, 787–799
38 Hamlin, J. L. (1992) *BioEssays* 14, 651–659
39 Benbow, R. M., Zhao, J. and Larson, D. D. (1992) *BioEssays* 14, 661–670
40 Fangman, W. L. and Brewer, B. J. (1992) *Cell* 71, 363–366
41 Kitsberg, D., Selig, S., Keshet, I. and Cedar, H. (1993) *Nature* 366, 588–590
42 Brown, W. R. A. (1992) *Curr. Opin. Genes Dev.* 2, 479–486
43 Farr, C., Fantes, J., Goodfellow, P. and Cooke, H. (1991) *Proc. Natl Acad. Sci. USA* 88, 7006–7010
44 Farr, C. J., Stevanovic, M., Thomson, E. J., Goodfellow, P. N. and Cooke, H. J. (1992) *Nature Genet.* 2, 275–282
45 Itzhaki, J. E., Barnett, M. A., MacCarthy, A. B., Buckle, V. J., Brown, W. R. A. and Porter, A. C. G. (1992) *Nature Genet.* 2, 283–287
46 te Reile, H., Maandag, E. R. and Burns, A. (1992) *Proc. Natl Acad. Sci. USA* 89, 5128–5132
47 Tyler-Smith, C. *et al.* (1993) *Nature Genet.* 5, 368–375
48 Larin, Z., Fricker, M. D. and Tyler-Smith, C. (1994) *Hum. Mol. Genet.* 3, 689–695
49 Curiel, D. T. *et al.* (1992) *Hum. Gene Ther.* 3, 147–154
50 Huxley, C. (1994) *Gene Ther.* 1, 7–12

120

EVIDENCE APPENDIX 3

U.S. Serial No. 10/001,486

In re Eynde, 480 F.2d 1364, 178 U.S.P.Q. 470 (CCPA 1973)

LEXSEE 480 F.2D 1364

IN THE MATTER OF THE APPLICATION OF HECTOR ALFONS VANDEN EYNDE, ROBERT JOSEPH POLLET AND ARTHUR HENRI DE CAT

Patent Appeal No. 8934

UNITED STATES COURT OF CUSTOMS AND PATENT APPEALS

480 F.2d 1364; 1973 CCPA LEXIS 307; 178 U.S.P.Q. (BNA) 470

July 19, 1973

PRIOR HISTORY: [**1] Serial No. 471,437.

CASE SUMMARY:

PROCEDURAL POSTURE: Appellants sought review of a decision of the patent office Board of Appeals sustaining the rejection of appellants' patent claims for chemical compounds based on a specification which failed to satisfy requirements of 35 U.S.C.S. §112.

OVERVIEW: Appellants' claims in a patent application for chemical compounds were rejected on the basis that the specification failed to satisfy the requirements of 35 U.S.C.S. §112. The patent examiner held the specification disclosure utility was insufficient to teach one skilled in the art how to use the claimed invention. The patent office Board of Appeals (board) sustained the rejection for lack of sufficient disclosure of how to use the compounds based on a new rationale, which involved comparing other patent not considered by the examiner. Appellants filed a request for reconsideration of board's decision, including a number of references to show the knowledge possessed by those skilled in the art, which the board refused to consider. The appellate court held board's refusal to consider evidence responding to its new rationale was in error and that it improperly regarded evidence on the use of appellants' compound.

OUTCOME: The court remanded the cause to the appeals board for further proceedings because it found the board erred in refusing to consider evidence bearing on its new rationale and by inquiring into the use of products obtained by reaction of appellants' claimed compounds.

CORE TERMS: compound, examiner, coupler, skilled, patent, specification, color, hydrazine, fluoroalkyl, invention, disclosure, photographic, pyrazoline, magenta, reconsideration, phenylhydrazine, possessed, pyrazolone, starting, phenyl, preparation, condensation, illustration, chemistry, chemist, colour, teach, atom, keto, aryl

LexisNexis(R) Headnotes

Patent Law > Anticipation & Novelty > General Overview
Patent Law > Jurisdiction & Review > Subject Matter Jurisdiction > Appeals
[HN1] An applicant is presumed to have all prior art before him at the time he makes an invention.

Patent Law > Claims & Specifications > Definiteness > General Overview
Patent Law > Claims & Specifications > Description Requirement > General Overview
Patent Law > Jurisdiction & Review > Subject Matter Jurisdiction > Appeals
[HN2] The first paragraph of 35 U.S.C.S. § 112 requires in pertinent part that: The specification shall contain a written description of the manner and process of using the invention in such full, clear, concise, and exact terms as to enable any person skilled in the art to use the same.

Patent Law > Claims & Specifications > Enablement Requirement > Standards & Tests
Patent Law > Jurisdiction & Review > Subject Matter Jurisdiction > Appeals
[HN3] The statutory requirement of 35 U.S.C.S. §112 is fulfilled where one possessed of the knowledge had by one skilled in the art could use the invention given the specification disclosure without undue experimentation.

A patent applicant may offer evidence, such as patents and publications, to show the knowledge possessed by those skilled in the art and thereby establish that a given specification disclosure is enabling. In such a situation, it is the knowledge possessed by those skilled in the art as of the filing date that is of relevance.

Patent Law > Claims & Specifications > Enablement Requirement > General Overview
Patent Law > Jurisdiction & Review > Subject Matter Jurisdiction > Appeals
Patent Law > U.S. Patent & Trademark Office Proceedings > General Overview
[HN4] When an examiner initially turns to an application, he may properly lodge a rejection of a claim as based upon a specification which is not in compliance with 35 U.S.C.S. § 112, first paragraph, if it appears reasonable to conclude that one skilled in the art would have been unable to use the invention at the time the application was filed. When that conclusion is reasonable, the burden is on to the applicant to rebut it, if he can, by offering evidence.

Patent Law > Claims & Specifications > Enablement Requirement > General Overview
Patent Law > Jurisdiction & Review > Subject Matter Jurisdiction > Appeals
Patent Law > U.S. Patent & Trademark Office Proceedings > General Overview
[HN5] If an applicant fails to challenge the examiner's conclusion regarding enablement, or fails to do so timely, the rejection stands.

Administrative Law > Agency Adjudication > Hearings > Evidence > General Overview
Patent Law > Jurisdiction & Review > Subject Matter Jurisdiction > Appeals
Patent Law > U.S. Patent & Trademark Office Proceedings > Examinations > General Overview
[HN6] It is the applicant's responsibility to prove that a prima facie reasonable conclusion of non-enablement is in fact unreasonable in view of the state of the art. Orderly and logical examining procedure compels that view. Moreover, the court rejects the notion that judicial or administrative notice may be taken of the state of the art. The facts constituting the state of the art are normally subject to the possibility of rational disagreement among reasonable men and are not amenable to the taking of such notice. If evidence of the knowledge possessed by those skilled in the art is to be properly considered, it must be timely injected into the proceedings.

Administrative Law > Agency Adjudication > Hearings > Evidence > General Overview
Patent Law > Jurisdiction & Review > Subject Matter Jurisdiction > Appeals
Patent Law > U.S. Patent & Trademark Office Proceedings > Examinations > General Overview
[HN7] Where the patent office Board of Appeals advances a position or rationale new to the proceedings, as it is empowered to do and quite capable of doing, the appellant must be afforded an opportunity to respond to that position or rationale by the submission of contradicting evidence.

OPINION BY:

LANE

OPINION: [*1366]

LANE, Judge.

This appeal is from the decision of the Patent Office Board of Appeals, adhered to on reconsideration, sustaining the rejection of claims 16-12 of appellants' application n1 as based on a specification which fails to satisfy the requirements of the first paragraph of 35 USC 112. We remand this case to the board for the reasons and purposes hereafter set forth. The most difficult aspect of this appeal is the determination of the propriety of the board's refusal to consider certain patents and publications submitted by appellants subsequent to the board's original decision as an accompaniment to a request for reconsideration. To fairly explain the issues before us and our resolution of them, we initially recount the positions of the principals as developed below.

n1 Serial No. 471,437 filed July 12, 1965, entitled "Fluoro-Alkylhydrazines and Process for the Preparation Thereof."

THE PROCEEDINGS BELOW

The Invention

The claims on appeal are all chemical compound claims. Claim 16 defines the genus and reads as follows:

16. A fluoro-alkyl hydrazine corresponding to the following formula:

[Graphic omitted. [**2] See illustration in original.]

wherein:

X is a member of the group consisting of a hydrogen atom and a fluorine atom,

R is a member of the group consisting of a hydrogen atom, lower alkyl or phenyl, and

n is a positive integer from 1 to 25 inclusive.

The specification characterizes the invention as relating to "new fluoro-alkyl-hydrazines" and "[more] particularly," "a process for the preparation thereof." Indeed, the specification for the most part discusses the manner of making the claimed compounds. As for utility, the specification states the following:

The fluoro-alkyl hydrazines according to the present invention are important starting materials for the synthesis of several organic compounds such as 1-fluoroalkyl-2-pyrazoline-5-one colour couplers, which form magenta azomethine dyes on colour development of exposed light-sensitive silver halide materials with aromatic primary amines.

Proceedings Before the Examiner

That statement of utility was the focus of the examiner's rejection of the claims first under both sections 101 and 112, and later under section 112 alone. The examiner considered the statement "vague and indefinite," and generally regarded [**3] it as insufficient to teach one skilled in the art how to use the claimed invention. The examiner questioned the manner in which the claimed hydrazines could be converted to color couplers and the manner in which the so-formed couplers could be used to yield magenta dyes on color development.

Appellants contended that one skilled in the relevant art would have been able to make color couplers from the hydrazines and would have known how to use the couplers in color development. Appellants generally argued that the utility requirements and all disclosure requirements of the patent laws were satisfied.

At one point during prosecution, appellants introduced a page of a treatise on photographic chemistry by Glafkide, [*1367] the relevant portion of which reads as follows:

> n2 Glafkide, Photographic Chemistry, vol. II, page 601, Fountain Press (London 1960).

(d) Pyrazolines. These magenta couplers with the structure

[Graphic omitted. See illustration in original.]

are very widely used for colour films. * * *

The pyrazolones are prepared by hot condensation of a phenylhydrazine Ar-NH-NH(2) with a keto compound such as acetoacetic ester CH(3)-CO-CH(2)-CO-O-C(2)H(5) [**4] in the presence of water and alcohol. They are insoluble in water, but soluble in caustic soda. Their sulphonated derivatives are soluble.

The simplest coupler is 1-phenyl -3-methyl -5-pyrazolone. The aryl sulphonated compound is prepared from sulphopheny lhydrazine. With diethyl-p-phenylene diamine, a magenta dye is obtained with the formula:

[Graphic omitted. See illustration in original.]

Before coupling, the Keto group -CO-CH(2)- is isomerized to the enol form -C(OH) CH-.

Appellants stated that:

[Glafkide] teaches the "classical" method of producing magenta color couplers with a pyrazolone structure by means of the hot condensation of a hydrazine with a carbonyl containing compound such as a substituted acetoacetic ester. Thus, the method of converting the fluoroalkyl hydrazine of the present invention to 1-fluoroalkyl-2-pyrazoline-5-one color couplers is well known in the art.

The examiner first commented on Glafkide in his

Answer stating:

The Glafkide publication does disclose the reaction of a phenylhydrazine with a keto compound to give a pyrazolone which then couples a primary aromatic amine to give a magenta azomethane compound, but the instant [**5] hydrazines are not phenyl hydrazines and there is no assurance that the instant hydrazines would so react or what the reaction conditions would be. [Emphasis in original.]

The examiner continued to hold that the specification disclosure of utility is insufficient, although he limited the basis of the rejection to the first paragraph of § 112.

In a reply brief following the Examiner's Answer, appellants expressly acknowledged the examiner's comment regarding the Glafkide publication.Appellants argued:

This statement would have greater significance if not for the expressed and implied indication by appellants that the conditions used with a phenylhydrazine when applied to the fluoroalkyl-hydrazines will result in a useful product.

Proceedings Before The Board

The board quoted the statement of utility appearing in appellants' specification and phrased the appellants' position as follows:

It appears to be appellants' view that, in the light of the Glafkide publication, one having ordinary skill in the art would know how to use the claimed compounds as

starting materials to produce the 1-fluoroalkyl-2-pyrazoline-5-one compounds, and furthermore would know how to employ [**6] the latter for the suggested purpose in photographic emulsions, all without further disclosure than that quoted above. [*1368]

The board agreed with the examiner's view of Glafkide stating:

We note * * * that the starting hydrazines and compounds produced by the Glafkide publication are different from those which would be employed and produced according to appellants' suggestion, in that they have aryl groups in place of the fluoroalkyl groups. We do not find that one skilled in the art would be able to predict with certainty that the fluoroalkyl compounds would react identically and under the same conditions as the aryl compounds of Glafkide, nor is there evidence in the record that they in fact do so.

In support of this conclusion the board made reference to a patent n3 (hereafter the Eynde patent) issued on an application n4 (hereafter the Eynde application) filed by appellants on July 14, 1965, two days subsequent to the filing date of the application involved in this appeal. The Eynde patent discloses in detail the manner of using the fluoroalkyl hydrazines to form color couplers and the use of such couplers in color development.

n3 U.S. Patent No. 3,462,270 issued August 19, 1969.

n4 Serial No. 472,017.

[**7]

It was presumably intended that the Eynde application be filed concurrently with the present application, and the present specification as originally filed made express reference to the Eynde application for a disclosure of the preparation of color couplers. However, the Eynde application was filed later for one reason or another, and reference thereto was deleted from the present application. The board was therefore aware of the Eynde patent which matured from the Eynde application and of the fact that appellants contemplated the methods disclosed therein as the means for converting the fluoroalkyl hydrazines to color couplers.

The board held as follows:

We note that in * * * [the Eynde patent] the conditions actually used are not the same as in the publication. The patent in all cases conducts the reaction in the presence of acetic acid; the reference employs a hot condensation in the presence of water and alcohol. It therefore is not even apparent from the record before us that the claimed compounds can in fact be used in the same manner as the known compounds of Glafkide * * *. Nor do we find that it would have been obvious that the fluoroalkyl pyrazoline-5-one compounds [**8] would be usable in the same manner in photographic compositions. * * * At the most, it might appear "obvious to try" the reactions and conditions indicated by the Glafkide publication upon appellants' novel hydrazine compounds, but this does not appear to be the standard by which sufficiency of disclosure under the first paragraph of 35 U.S.C. 112 is determined. We will therefore sustain the rejection for lack of sufficient disclosure of how to use the compounds.

We observe at this point that the examiner had not based his conclusions on the differences between the Eynde patent reaction environment and that disclosed in Glafkide. That was clearly a rationale new to the proceedings. We cannot ascertain the extent to which the board's rationale influenced its affirmance. For example, we do not know whether the board considered it an alternative basis for affirmance or whether the board regarded it is providing essential support to the examiner's reasoning.

Appellants filed a request for reconsideration of the decision of the board accompanied by a number of patent and publication references offered for the purpose of demonstrating the knowledge possessed by those skilled in [**9] the art. The board noted that the references were not "urged upon [it] in the brief," and since the references were not of record prior to the board's original decision, the board refused to consider them. [*1369] The board adhered to its original decision.

In response to the board's refusal to consider the reference material, appellants contended that the board erred in its assessment of appellants' position, noted above, to the effect that "in the light of the Glafkide publication," one skilled in the art would have known how to use the claimed compounds. Appellants justified the submission of additional reference material as follows:

Appellants asserted and assert that having the teaching of the specification, one skilled in the art of photographic chemistry would be able to employ the invention in light of the disclosure of utility at page 3 of appellants' specification. The Glafkide reference was referred to as being exemplary of what was known and available to the skilled photographic chemist. It was not intended to be a summary of the only knowledge known and available to the skilled photographic chemist. Appellants, in their Petition for Reconsideration, referred [**10] to the numerous other patents and [publications] to clearly establish the host of art in this area. This is art which was available to the Examiner in the Patent Office and to the Board of Appeals to determine what knowledge was available to a skilled photographic chemist. This art could have been obtained merely by calling an examiner

in Group 160, the art unit responsible for examining photographic cases.[HN1]

An applicant is presumed to have all prior art before him at the time he makes an invention. It is submitted that it is not too much for an applicant to ask that the examiner and the Board of Appeals be held to the same standard, with it being presumed that all the pertinent art was available to the Examiner. [Emphasis in original.]

The Positions of the Parties on Appeal

The solicitor states that the rejection here involved is based on the "how-to-use" requirement of the first paragraph of section 112. It is emphasized to the court that the truth of the assertions that the fluoroalkyl hydrazines can be converted to pyrazoline-one color couplers and that such color couplers function in the manner described in the specification is not here questioned. Rather, the gist [**11] of the rejection is said to be that one skilled in the art would not have known how to so utilize the claimed compounds, at least in the absence of probative evidence to the contrary.

The solicitor urges that the board correctly refused to consider the reference material submitted after its decision. He observes that the board agreed with the examiner that Glafkide is insufficient since directed to phenylhydrazines, and he argues that none of the references subsequently filed bears directly on the additional point raised by the board respecting the differences between the Glafkide and Eynde patent reaction environment conditions. The solicitor concludes that the board made no new factual assumptions the rebuttal of which might justify considering additional reference material.

Appellants contend that the claimed compounds have utility and that one skilled in the art would know how to use them given the specification statement of utility. Appellants additionally contend that Glafkide demonstrates the manner in which the claimed compounds can be used. They argue that as a matter of fact, the phenyl portion of the Glafkide structure, like the fluoroalkyl group of their own compound, [**12] is not involved in the pyrazoline-one-forming reaction.

Appellants assert that in any event the additional reference material clearly establishes that at the time the present application was filed, the level of skill in the pertinent art was such that the manner of using the compounds would have been apparent. It is appellants' position that a "new rationale" of the board prompted the submission of that material and that In re Moore, 58 CCPA 1340, 444 F.2d 572, 170 USPQ 260 (1971) is authority for the proposition that the [*1370] board should have considered it. Alternatively, they ask this court to take judicial notice of those prior patents and publications.

OPINION

The "How-to-Use" Requirement

We accept the solicitor's view that the rejection before us is based solely on the "how-to-use" requirement of § 112. [HN2] The first paragraph of § 112 requires in pertinent part that:

The specification shall contain a written description * * * of the manner and process of * * * using * * * [the invention] in such full, clear, concise, and exact terms as to enable any person skilled in the art * * * to * * * use the same * * *.[HN3]

That statutory requirement is fulfilled where [**13] one possessed of the knowledge had by one skilled in the art could use the invention given the specification disclosure without undue experimentation. A patent applicant may offer evidence, such as patents and publications, to show the knowledge possessed by those skilled in the art and thereby establish that a given specification disclosure is enabling. See, e.g., Martin v. Johnson, 59 CCPA , 454 F.2d 746, 172 USPQ 391 (1972). In such a situation, it is the knowledge possessed by those skilled in the art as of the filing date that is of relevance. See Tummers v. Kleimack, 59 CCPA , , 455 F.2d 566, 568, 172 USPQ 592, 593 (1972).

[HN4] When an examiner initially turns to an application, he may properly lodge a rejection of a claim as based upon a specification which is not in compliance with § 112, first paragraph, if it appears reasonable to conclude that one skilled in the art would have been unable to use the invention at the time the application was filed.When that conclusion is reasonable, the burden is on to the applicant to rebut it, if he can, such as by offering evidence as discussed above. See, e.g., In re Gardner, 475 F.2d 1389, 177 USPQ 396 (CCPA 1973).

It is clear that [**14] an applicant must be permitted to respond to a position assumed by the examiner or the board. However, the response must be timely. [HN5] If an applicant fails to challenge the examiner's conclusion regarding enablement, or fails to do so timely, the rejection stands.

We reject the contention advanced by appellants that evidence submitted by an applicant in response to a § 112 rejection is timely at any point during prosecution because the examiner or board could have located it. [HN6] We adhere to the view that it is the applicant's responsibility to prove that a prima facie reasonable conclusion of nonenablement is in fact unreasonable in view of the state of the art. Orderly and logical examining procedure compels that view. Moreover, we reject the notion that judicial or administrative notice may be taken of the state of the art. The facts constituting the state of the art are normally subject to the possibility of rational

disagreement among reasonable men and are not amenable to the taking of such notice. If evidence of the knowledge possessed by those skilled in the art is to be properly considered, it must be timely injected into the proceedings.[HN7]

We do agree with appellants [**15] that where the board advances a position or rationale new to the proceedings, as it is empowered to do and quite capable of doing, the appellant must be afforded an opportunity to respond to that position or rationale by the submission of contradicting evidence. This court so held in In re Moore, supra, and we expressly reaffirm that view. The board's refusal to consider evidence which responds to such a new rationale is error.

The Board's Refusal to Consider the Evidence in This Case

Appellants in this case strenuously and primarily asserted that the board misconstrued the import of the submission of the Glafkide reference during prosecution before the examiner. They [*1371] argued that the board treated Glafkide as the only evidence of the state of the art whereas they intended it as merely exemplary of many patents and publications in the field of photographic chemistry. In appellants' view, the position of the board as to the role of Glafkide in establishing the state of the art was itself a new rationale which justified the submission of additional evidence.

Appellants' contention is untenable. The board originally decided the appeal on the record before it. [**16] Glafkide was the only reference material which appellants had offered to that point.

In discussing the submitted material in the request for reconsideration, appellants pointed to disclosures which allegedly tend to support the view that those skilled in the art would have considered the reaction of a phenylhydrazine to form a pyrazoline color coupler to be directly applicable to the conversion of a fluoroalkyl hydrazine contrary to the examiner's and board's belief. Appellants were confronted with this argument in the Examiner's Answer and expressly acknowledged it, and attempted to answer it, in a reply brief. If there is evidence to support their position, appellants should have presented it at that point. The board did not raise that issue, and we agree with the solicitor that there is no basis in the board's decision for a justification for the submission of evidence on that issue after the board's decision. The board did not err in refusing to consider the later-filed evidence to that extent.

However, the board clearly went off on its own in considering the differences between the Glafkide reaction environment and the Eynde patent reaction environment. Appellants [**17] offered evidence allegedly showing the prior art use of the Eynde patent conditions in hydrazine-to-color coupler reactions. By demonstrating that both the Glafkide and Eynde patent reaction conditions were known prior to filing, appellants might have convinced the board that at the time of filing, one skilled in the art would not have had reason to doubt the efficacy of the Glafkide environment. We disagree with the solicitor's analysis of the evidence offered on this point, and we conclude that the board did commit error in refusing to consider that evidence in rebuttal of the board's newly advanced position.

The Manner of Using the Color Couplers

Thus far in our opinion we have not mentioned the board's findings with respect to the manner of using the color couplers. We agree that in order for the specification to be enabling in the sense of § 112, one skilled in the art would have to have been able to use the claimed hydrazines in the manner asserted in the specification, i.e. as a starting material for the synthesis of color couplers. However, color couplers are, in general, unquestionably well known materials, and utility for color couplers per se need not be separately [**18] established. We accordingly see no necessity for appellants to demonstrate how to use the color couplers. Given that the claimed compounds are useful in the manufacture of products having known utility, appellants' specification is sufficient if it enables one skilled in the art to make those products. We accordingly hold that the board was also in error to the extent that it agreed with the examiner that the specification is insufficient for failure to teach how to use the phyrazoline-one color couplers.

The Remand

We have made two findings of error which seriously undermine the decision of the board. We have found that the board erroneously refused to consider evidence allegedly bearing on its new rationale and that it erroneously regarded inquiry into the manner of using products obtained by reaction of the claimed compounds as proper. We remand this case to the board for action not inconsistent with these findings or this opinion in general.

REMANDED

EVIDENCE APPENDIX 4

U.S. Serial No. 10/001,486

Ex parte Forman, et al., 230 U.S.P.Q. 546 (BPAI 1986)

230 U.S.P.Q. 546

1986 WL 83597 (Bd.Pat.App. & Interf.), 230 U.S.P.Q. 546

(Cite as: 230 U.S.P.Q. 546)

C

Ex parte Forman, et al.

Patent and Trademark Office Board of Patent Appeals and Interferences

Appeal No. 602-90

Decided February 5, 1986 and April 22, 1986
Released July 2, 1986

United States Patents Quarterly Headnotes

PATENTS
[1] Specification -- Sufficiency of disclosure (§ 62.7)
Impossibility of determining, from reading of specification and record, number of mutant strains of original S. typhi and hyperconjugant strains of genetically engineered hybrid that are originally formed in each experiment, nor what amount of time, effort, and level of skill is needed to isolate single strains which can then be cloned to yield useful vaccines, supports conclusion that practice of invention would require undue experimentation.

PATENTS
Particular Patents -- Vaccine
Forman, Baron, and Kopecko, application No. 289,013, Oral Vaccine for Immunization Against Enteric Disease, rejection of claims 1-33 affirmed.

***546** Appeal from Art Unit 127.

Application for patent of Samuel B. Forman, Louis S. Baron, and Dennis J. Kopecko, application, Serial No. 289,013, filed July 31, 1981, for Oral Vaccine for Immunization Against Enteric Disease. From decision rejecting claims 1-33, applicants appeal. Affirmed.

William G. Gapcynski, Werten F.W. Bellamy, Francis A. Cooch and John H. Raubitschek, all of Washington, D.C., for appellants.

Before Serota, Milestone, and Goldstein, Examiners-in-Chief.

Goldstein, Examiner-in-Chief.

This appeal is from the examiner's final rejection of claims 1 to 33. Claims 34 to 40 have been allowed. Illustrative rejected claim 1 and allowed claim 34 are reproduced below.

> 1. A living attenuated oral vaccine, for the immunization against enteric disease, comprising as the active component an effective dose of a genetic hybrid derivative of an attenuated galactose epimeraseless mutant strain of S. typhi and at least one non-typhoid protective antigen carried thereby, ***547** or a mixture of said genetic hybrid derivatives.
>
> 34. A genetic hybrid derivative, having deposit accession number ATTC 31904, of an attenuated galactose epimeraseless mutant strain of S. typhi and a non-typhoid protective antigen carried thereby, wherein the non-typhoid protection antigen is the form I antigen of Shigella sonnei and the genetic hybrid derivative strain expresses both S. typhi and S. sonnei antigens.

No references have been relied on by the examiner on appeal. References incorporated into the disclosure of the present specification by appellants are:

Germanier 3,856,935 Dec. 24, 1972

Kopecko, D.J. et al. (Kopecko), Infect. Immun. 29:207-214, 1980.

Claims 1 to 33 have been finally rejected under the first paragraph of 35 U.S.C. 112 as being based on an insufficiently enabling disclosure. We shall affirm this rejection.

Appellants' claimed invention is a class of oral vaccines consisting of hybrid bacteria which have been genetically engineered (specification, page 6, ultimate line) to produce an immunizing effect against both the bacterial antigen (typhoid) and one other non-typhoid enteric disease. A commercially available Salmonella typhi strain is converted into streptomycin resistant mutants by mutagenesis (specification, page 9, penultimate paragraph). The antigen for the non-typhoid disease is disclosed as being introduced into the mutated S. typhi strain through a variety of techniques utilizing both plasmid-borne genetic determinants and bacterial chromosomes (specification, page 6, last sentence of first paragraph). There is, however, no discussion in any detail whatever as to anything but the plasmid-borne embodiment. This embodiment involves introduction of the plasmid-borne antigen of the non-typhoid disease from a variety of mutant strains of enteric bacteria, such as, for example, Shigella, through a technique known as transconjugation (see paragraph bridging pages 7 and 8 of the specification, for example).

The basis of the examiner's rejection appears to be generally two-fold. It appears to be the examiner's position that, in the absence of a deposit, one of ordinary skill in the relevant art would not be able to obtain the specific S. typhi mutant apparently essential to appellants' invention. See specification page 20, lines 11-14. It also appears to be the examiner's position that the hyperconjugation procedure is not sufficiently well known and straightforward so that one of ordinary skill in this art could predict the results with an adequate degree of certainty based on only the record before her to justify the allowance of generic claims as opposed to those claims based on available, deposited, specific strains. For the reasons which we shall elaborate below, we conclude that on the record before us the examiner's position is correct.

The ultimate question in each case of this type is whether or not the specification contains a sufficiently explicit disclosure to enable one having ordinary skill in the relevant field to practice the invention claimed therein without the exercise of undue experimentation. The requirement for enablement can be found expressly stated in the first paragraph of 35 U.S.C. 112, which requires that the disclosure of an invention be "in such full, clear, concise and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same . . . ." The "undue experimentation" proscription is, in effect, a gloss on the statute which has arisen from decisional law which requires that sufficient information be given in the application so that one of ordinary skill in the art can practice it without the necessity for undue experimentation: Fields v. Conover, 58 CCPA 1366, 443 F.2d 1386, 170 USPQ 276 (1971); In re Colianni, 561 F.2d 220, 195 USPQ 150 (CCPA 1977). Also see Ansul Co. v. Uniroyal, Inc., 448 F.2d 872, 169 USPQ 759 (2d. Cir. 1971), cert. denied, 404 U.S. 1018, 172 USPQ 257 (1972). The principle that the disclosure of a patent application must enable practice of the invention claimed without undue experimentation was well established long before the enactment of the 1952 Patent Act. See Standard Brands, Inc. v. National Grain Yeast Corp., 101 F.2d 814, 40 USPQ 318 (3rd Cir. 1939), and cases cited therein.

The determination of what constitutes undue experimentation in a given case requires the application of a standard of reasonableness, having due regard for the nature of the invention and the state of the art: Ansul Co. v. Uniroyal, Inc., supra. The test is not merely quantitative, since a considerable amount of experimentation is

permissible, if it is merely routine, or if the specification in question provides a reasonable amount of guidance with respect to the direction in which the experimentation should proceed to enable the determination of how to practice a desired embodiment of the invention claimed. The factors to be considered have been summarized as the quantity of experimentation necessary, the amount of direction or guidance presented, the presence or absence of working examples, the nature of the invention, the state of the prior art, the relative skill of those in that art, the predictability or unpredictability of the art and the breadth of the claims. In re Rainer, 52 CCPA 1593, 347 F.2d 574, 146 USPQ 218 (1965); In re Colianni, supra.

***548** The examiner in this case has relief on the published decision in Ex parte Jackson, 217 USPQ 804 (PTO Bd. App. 1982), as supporting her position. Appellants on the other hand have pointed out the distinction that the issue in the Jackson case was the undue experimentation involved in finding additional strains over and above the deposited cultures in nature, whereas here the relevant hybrid bacterial strains are produced by genetic engineering by processes which at least begin with available bacterial strains. Although we recognize the distinction noted by appellants, we find that there is insufficient evidence of record in the present case to suggest that the burden of experimentation on one of ordinary skill in the relevant art would be any less in this case.

[1] We recognize that the level of skill in the art of molecular biology is quite high. Nonetheless, we also recognize the correctness of the examiner's position that experiments in genetic engineering produce, at best, unpredictable results. From a reading of appellants' specification and the Kopecko article of record, [FNa1] it is impossible to determine how many mutant strains of the original S. typhi and hyperconjugant strains of the genetically engineered hybrid are originally formed in each experiment and how much time and effort and what level of skill must be exercised to isolate the single strains which are then cloned to yield the useful vaccines. If there are any known clues to assist one of ordinary skill in predicting which of the myriad strains that are presumably produced would be useful, there is nothing on the present record to suggest what they might be. We note that appellants have acknowledged (Paper No. 6, top of page 12) that "the genetic techniques involved are very time consuming" and that it is anticipated that it "will take about one year to construct" most strains according to this invention. It is readily apparent from our above discussion of the relevant law that time is not the sole criterion. However, in the present case there is also a lack of guidance leading to predictability as discussed in the immediately preceding portion of this opinion. The art under consideration appears to be relatively undeveloped. There do not appear to be any apparently precisely reproducible working examples, and most of the claims are broad in scope. Consideration of all of these factors leads to the conclusion that practice of the rejected claims would require undue experimentation. In re Rainer, supra; In re Colianni, supra.

As we have noted above, there does not appear to be on the present record, even including the Kopecko article, a single detailed example which could be followed by another worker in another lab to obtain a single specific microorganism (vaccine) within appellants' claims, without recourse to the deposited strains recited in the allowed claims. Appellants have requested that we indicate what microorganism deposits and supporting documentation would be required to result in the allowance of broader claims (see Reply to the Supplemental Examiner's Answer, Paper No. 20). Whether or not it would be proper for us to make such a determination in any event, it should be clear from our above discussion of the present record that we find it inadequate to enable us to respond even speculatively to appellants' request.

The decision of the examiner is affirmed.

AFFIRMED

ON REQUEST FOR RECONSIDERATION

This is a request for reconsideration of our

decision of February 5, 1986, wherein we affirmed the examiner's rejection of claims 1 to 33.

We have carefully considered the arguments advanced by appellants but find nothing therein to convince us that the decision was in error. We have carefully considered particularly the section of the specification bridging pages 8 and 9, to which appellants referred in the request for reconsideration.

We note that construction of the donor strain involved four consecutive mating or transconjugation procedures. This procedure broadly may be well known, but there is nothing in evidence to indicate how complex the results might be in this particular case. Selection was required after each consecutive procedure, and there is no indication as to the frequency of the desired transfer in a cell population from which one could judge how much trial and error was involved in the selection.

We believe the examiner's characterization and ours of the procedures disclosed as being unpredictable was correct on this record. Appellants have added nothing to the record to convince us to the contrary but merely repeated arguments similar to those already presented before the examiner. There is no evidence properly before us to establish that one of ordinary skill in this art has actually been able to produce additional species representative of any of the claims broader than those allowed ***549** by the examiner. We emphasize that the declaration of Dennis J. Kopecko, filed under 37 C.F.R. 1.132, was refused consideration by the examiner as being untimely. Thus, it is not before us, and it would be inappropriate for us even to speculate upon its potential probative value.

It should be emphasized that the absence of a working example was not in itself stated to cause the specification to be insufficiently enabling. It was mentioned merely as one of the factors set forth in In re Rainer, 52 CCPA 1593, 347 F.2d 574, 146 USPQ 218 (1965), and In re Colianni, 561 F.2d 220, 195 USPQ 150 (CCPA 1977). Since there appears to be no direct control over the various mating and transconjugation procedures set forth in the specification (see also those pages following page 9, where additional selection is indicated) it may, in fact, be impossible to present an exactly reproducible working example in cases of this type. If true, however, this fact would merely further emphasize the requirement for appropriate deposits of useful organisms to enable the reasonably ready practice of inventions of this type.

The request has been considered but is denied with respect to making any change in our decision.

DENIED

FNa1 The additional literature citation (Hayes) and the Kopecko declaration under 37 CFR 1.132, to which it was attached, were considered untimely filed by the examiner and thus are not part of the record considered by us.

P.T.O. Bd.Pat.App. & Int.

230 U.S.P.Q. 546

END OF DOCUMENT

L. RELATED PROCEEDINGS APPENDIX

None